Contract No. W911NF-07-D-0001 / TCN 08319/DO 0567

REPORT

Laboratory Medicine Best Practices: Developing Systematic Evidence Review and Evaluation Methods for Quality Improvement Phase 3 Final Technical Report

Prepared for:

Division of Laboratory Science and Standards

Laboratory Science, Policy and Practice Program Office
Office of Surveillance, Epidemiology, and Laboratory Sciences (OSELS)
Centers for Disease Control and Prevention

*Prepared by the CDC Laboratory Medicine Best Practices Team**

May 27, 2010

** Susan Snyder, Edward Liebow, Colleen Shaw, Robert Black, Robert Christenson, James Derzon, Paul Epner, Alessandra Favoretto, Lisa John, Diana Mass, Abrienne Patta, Shyanika Rose, Malaika Washington*

Laboratory Medicine Best Practices: Developing Systematic Evidence Review and Evaluation Methods for Quality Improvement Phase 3 Final Technical Report

EXECUTIVE SUMMARY

BACKGROUND AND PURPOSE: This report summarizes the third phase of an ongoing effort sponsored by the Division of Laboratory Science and Standards (DLSS), Centers for Disease Control and Prevention. The purpose is to develop new systematic evidence review and evaluation methods for identifying pre- and post-analytic laboratory medicine practices that are effective at improving healthcare quality.[1] This effort began in 2006, when CDC convened the Laboratory Medicine Best Practices Workgroup (Workgroup), a multidisciplinary panel of experts in such fields as laboratory medicine, clinical medicine, health services research, and health care performance measurement. The Workgroup also includes two ex officio representatives from the Centers for Medicare and Medicaid Services (CMS) and the Food and Drug Administration (FDA).

An outcome of Phase 1 (2006 – 2007) was to act on a Workgroup recommendation to enlarge the search for evidence to unpublished studies, including assessments performed for the purposes of quality assurance, process improvement and/or accreditation documentation. Phase 2 (2007-2008) involved a pilot test of further refined methods to obtain, review, and evaluate published and unpublished evidence, along with collecting observations via key informant interviews about organizational and implementation issues successfully addressed by other recommending bodies about the development and dissemination of guidelines and best practice recommendations. These evidence review methods were adapted from those established by the GRADE group, The Guide to Community Preventive Services (Community Guide), the Agency for Healthcare Research and Quality (AHRQ) (US Preventive Services Task Force (USPSTF), Evidence-based Practice Centers (EPCs), and Effective Healthcare Program), and others, and modified to better accommodate the non-controlled study designs typically found in quality improvement research.

Phase 3 (2008-2010), the subject of this report, involved further development of methods for identifying evidence-based laboratory medicine quality improvement best practices, and validated these methods with reviews of practices associated with three topics: patient specimen identification, critical value reporting, and reducing blood culture contamination.

SYSTEMATIC EVIDENCE REVIEW AND EVALUATION METHODS: Methods developed in earlier phases were refined and applied to identify and frame review topics and questions, and then collect, screen, abstract, standardize, summarize, and evaluate

[1] The LMBP Initiative relies on the Institute of Medicine's six healthcare quality domains of safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity for measuring and evaluating laboratory medicine practice effectiveness (Committee on the National Quality Report on Health Care Delivery, 2001).

evidence from published and unpublished sources for specific practices/interventions. The approach to implementing these evidence review steps adopted the vocabulary of a framework commonly used in evidence-based medicine (Ask-Acquire-Appraise-Analyze-Apply-Assess, or "A-6", (Shaneyfelt et al 2006)). These methods include the guidance provided to expert panelists, who were asked to (1) review and finalize study quality ratings drafted by the review team; (2) evaluate and rate the magnitude of effect sizes obtained from these studies and their consistency; (3) use these ratings to assess the overall strength of a body of evidence for a given practice; (4) present their evaluation findings; and then (5) translate their findings for each practice into a draft evidence-based recommendation.

The expert panels' evidence reviews, evaluations, and draft recommendations became the basis for consideration of best practice recommendations by the Workgroup (serving in its capacity as the "Recommending Body"). As with earlier phases, methods for including rating and evaluating study findings for a practice-specific evidence base were adapted from protocols from several organizations involved with public health and healthcare-related evidence reviews and recommendations.[2]

A key Phase 3 objective was to examine the utility and feasibility of including unpublished assessments or studies as part of the systematic evidence reviews of laboratory medicine practices (LMBP). Established steps for collecting evidence from unpublished sources included:

1. Obtaining the support and endorsement of key stakeholder organizations to encourage clinical laboratories and healthcare organizations to participate in the LMBP pilot test.

2. Identifying healthcare organizations/facilities likely to have completed relevant unpublished laboratory medicine practice assessments, based on:
   a. Conference papers or other public presentations.
   b. Relevant publications that implied the author(s) or others might have additional data beyond what was reported (e.g., more recent data, or data more encompassing in scope or care setting)
   c. Personal knowledge of Workgroup and Expert Panel members and the CDC/Battelle team.
   d. Calling attention to an online site where facilities could voluntarily register their interest in being contacted to gauge whether available data would be appropriate for inclusion.

3. Identifying and contacting a senior laboratory scientist, laboratory director, or other appropriate representatives (e.g., involved in patient safety, quality management, clinical research, regulatory/accreditation compliance) to

[2] The Guide to Community Preventive Services (http://www.thecommunityguide.org/index.html), the US Preventive Services Task Force (http://www.ahrq.gov/clinic/uspstfix.htm), The GRADE Working Group (http://www.gradeworkinggroup.org/index.htm), AHRQ (EPCs http://www.ahrq.gov/Clinic/epcpartner/epcresmat.htm and Effective Healthcare Program http://www.effectivehealthcare.ahrq.gov/index.cfm/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productid=318) The Cochrane Collaboration (http://www.cochrane.org/).

describe the aims of the LMBP project and explore the circumstances under which the organization would consider participating in the pilot test.

4. Providing additional information about the project to the facility point-of-contact to share with colleagues and obtain a preliminary assessment from the organization's Institutional Review Board (IRB) chair for release of de-identified data from previously completed studies What was found early on at the end of the first year of this initiative is that we weren't going to be able to adopt conventional ways for doing evidence reviews to laboratory medicine quality improvement practices due to generally insufficient published evidence. There was a recognition that Available evidence/data was likely to come from quality improvement efforts which don't tend to get published.

5. Extending a formal invitation to the organization and providing more general guidance about the type of information needed from unpublished studies.

6. Establishing any formal confidentiality safeguards or conditions under which the information would be provided for the purposes of the pilot test of LMBP systematic review methods.

7. Reviewing study information and other material received, and follow-up with additional information requests as needed.

To minimize the burden on pilot test participants and maintain consistency with published evidence, only previously completed studies were requested (i.e., no new data), and it was suggested that these studies might be derived from multiple types of sources, including internal assessments, case studies, Failure Mode and Effects Analyses (FMEA), and quality improvement project studies. Facilities were also requested to provide data that contained no personal patient health information. A commitment was made to de-identify all data and studies submitted, and each facility was offered the option to remain anonymous in the pilot test evidence summaries and findings.

All studies and/or assessments, published and unpublished, acquired for the pilot LMBP evidence reviews were screened using the same criteria for relevance and completeness (i.e., had at least one effectiveness finding for a practice being reviewed with an outcome measure associated with the review question). Studies that met the inclusion criteria were then abstracted by at least two independent reviewers, summarized in a standardized format, and included in evidence summaries and meta-analyses for each practice reviewed. The evidence summaries and LMBP study quality rating criteria were used to categorically rate individual study quality, and the individual study and meta-analysis summary effect sizes were also categorically rated to produce an overall strength of evidence rating for each practice, using the following four-step approach:

1. Categorically rating individual study quality (good, fair, poor), based on a 10-point scale with specified criteria evaluating four quality dimensions
    a. Study
    b. Practice
    c. Outcome measure(s)
    d. Findings/result(s)

2. Categorically rating the observed effect size(s) (substantial, moderate, minimal/none) reported in each individual study with a "good" or "fair" study quality rating and relevance to the review question (direct, less direct, indirect). (Studies with "poor" quality ratings are excluded from the practice evidence base and the effect-size meta-analyses).

3. Assessing the consistency of all study effect sizes based on their direction and magnitude.

4. Rating the overall strength of a body of evidence using the ratings from the three previous steps is based on the number of good and fair quality studies that found a substantial or moderate effect size.

The following are the established rating categories for the overall strength of a body of evidence:

High: An adequate volume of evidence is available and includes consistent evidence of substantial healthcare quality impact from studies without major limitations.

Moderate: Some evidence is available and includes consistent evidence of substantial healthcare quality impact from studies without major limitations; OR an adequate volume of evidence is available and includes consistent evidence of moderate healthcare quality impact from studies without major limitations.

Suggestive: Limited evidence is available and includes consistent evidence of moderate healthcare quality impact from a small number of studies without major limitations; or the quality of some of the studies' design and/or conduct is limited.

Insufficient: Any estimate of an effect on healthcare quality impact is too uncertain. Available evidence of effectiveness is:

- Inconsistent or weak; OR
- Consistent but with a minimal effect; OR
- Contained in an inadequate volume to determine effectiveness

EVIDENCE-BASED IDENTIFICATION OF BEST PRACTICES: The rating categories for the overall strength of a body of evidence related to a potential best practice translates into recommendation rating categories. These rating categories reflect the extent to which there is confidence that the available evidence demonstrates that the practice(s) will do more good than harm:

Recommend: The practice should be identified as a "best practice" for implementation in appropriate care settings, taking into account variations and applicability in implementation and/or care settings. This recommendation results from a "High" or "Moderate" overall strength of evidence rating for improving healthcare quality, and accounts for available information related to additional harms and benefits.

No recommendation for or against: A potentially favorable impact on healthcare quality is not of sufficient size, or not sufficiently supported by evidence to indicate that it should be identified as a "best practice" for

implementation in appropriate care settings. This recommendation results from a “Suggestive” or “Insufficient” overall strength of evidence rating, and accounts for available information related to additional harms and benefits.

Recommend against: The practice should not be identified as a “best practice” for implementation because it is not likely to result in more good than harm. This recommendation results from a “High” or “Moderate” overall strength of evidence rating for adversely affecting healthcare quality, and accounts for available information related to additional harms and benefits.

There is an important distinction between evidence of effectiveness for healthcare quality improvement and evidence related to other aspects of implementation, such as feasibility, cost, applicability (e.g., to specific care settings and populations), and other harms and benefits. Only the evidence of effectiveness was systematically reviewed. Further methods refinements for these implementation aspects will be considered in future reviews.

PHASE 3 EVIDENCE REVIEW RESULTS: Seven practices met the pilot test minimum criteria for available evidence to be considered for systematic reviews: two for the Patient Specimen Identification topic, two for the Communicating Critical Values topic, and three for the Blood Culture Contamination topic.

Patient Specimen Identification: Practices associated with this review topic are designed to reduce patient specimen and/or test result identification errors and assure accurate identification of specimens and/or test results. Practices for which enough evidence was available from unpublished and published sources to be included in the evidence review were:

- **Barcoding Systems** - Electronic bar-coding of both patient identification and specimen used to establish positive identification of specimen as belonging to patient. This involves the use of bar code scanners and capability to barcode specimens.
- **Point-of-Care-Testing Barcoding Systems** - Automated patient and sample/test result identification system using bar-coded patient identification and bar code scanners when using a testing device at or close to the patient.

Critical Values Communication: Practices associated with this review topic are designed to assure timely and accurate communication of critical value laboratory test results to a licensed responsible caregiver who can act on these results. Practices for which enough evidence was available from unpublished and published sources to be included in the evidence review were:

- **Automated Notification** – Automated alerting system or computerized reminders using mobile phones, pagers, email or other personal electronic devices to alert clinicians of critical value laboratory test results.
- **Call Center** – Critical value notification process centralized in a unit responsible for communication of critical value laboratory test results to the licensed caregiver.

Blood Culture Contamination: Practices associated with this review topic are designed to reduce blood culture contamination rates (i.e., false positive blood culture test results

associated with contaminants in blood culture specimens), which routinely result in unnecessary repeat tests and antimicrobial drug therapy associated with adverse clinical and economic outcomes (e.g., increased hospital length of stay, side effects, and cost of therapy). Practices for which enough evidence was available from unpublished and published sources to be included in the evidence review were:

- **Dedicated Phlebotomy** – Use of certified phlebotomists (rather than nursing or other staff) to draw blood specimens for analysis, acknowledging that 100% of phlebotomist blood draws use venipuncture collection.
- **Venipuncture (vs. Intravenous catheter) collection** – Puncture of a vein through the skin *vs.* use of a thin flexible tube inserted into the body to withdraw blood for analysis
- **Pre-packaged Prep Kits** - Pre-packaged aseptic supplies for drawing blood specimens by venipuncture that are prepared in-house or commercially purchased

Preliminary results (December 2009): Based on the strength of evidence, the following were identified as "best practice" recommendations.

Patient Specimen Identification:

- The use of barcoding systems (vs. no barcoding) is identified as a best practice for reducing patient specimen identification errors (8 studies, log odds ratio = 2.45; 95% CI 1.6-3.3).
- The use of point-of-care-testing barcoding systems is identified as a best practice for reducing patient test result identification errors (5 studies, odds ratio 6.55; 95% CI 3.1 – 14.0).

Critical Value Reporting:

- No recommendation is made for or against identifying the use of call centers (3 studies, Standard difference of means = 0.81, 95% CI -0.52 – 2.15) [3] or automated notification systems (3 studies, Standard difference of means = 0.51, 95% CI -0.4 – 1.4) as a best practice.

Blood Culture Contamination:

- The use of venipuncture for sample collection when this option exists in the clinical setting is identified as a best practice for reducing blood culture contamination rates (7 studies, OR = 2.63, 95% CI 1.85-3.72).
- The use of dedicated phlebotomy (teams) to collect blood culture specimens is identified as a best practice for reducing blood culture contamination rates (6 studies, OR = 2.76, 95% CI 2.2 - 3.5).

[3] When the Confidence Interval (CI) for the Odds Ratio extends below 1.0 (or below 0.0 for the Standard Difference of Means), we cannot determine whether there is an effect that favors the intervention over the comparator.

- No recommendation is made for or against identifying the use of pre-packaged preparation kits (4 studies, OR =1.1, 95% CI 0.99-1.41)[3] as a best practice.

CONCLUSIONS

Methods

- Findings from pilot LMBP systematic reviews (2006-2009), demonstrate that LMBP systematic review and evaluation methods may be applied to evaluate quality improvement practices.
- Systematic evidence review and evaluation methods developed and tested during Phase 2 were refined and adapted to better address the evidence available from laboratory medicine quality improvement studies resulting in greater consistency and transparency of evidence rating and evidence.
- Unpublished and published data from laboratory quality improvement efforts provide evidence of effectiveness for inclusion in systematic evidence reviews.
- The Phase 3 pilot test findings demonstrate that LMBP systematic review methods for quality improvement practice evidence reviews support evidence-based recommendations. The LMBP methods for summarizing and evaluating practice evidence of effectiveness, and rating the overall strength of a body of evidence are comprehensive, appropriate and can be efficiently implemented on an ongoing basis given sufficient organizational resources and appropriately qualified staff, but still require further specific refinements in Phase 4 (ending in 2011) discussed below.

Network for unpublished evidence

Phase 3 efforts to recruit healthcare organizations to participate in a network to provide unpublished evidence provided considerable insight into the factors that constrain and encourage participation, and the likelihood of obtaining usable evidence, including:

- Contacts with knowledgeable representatives invested with appropriate decision-making authority,
- Identification and participation of organizations that use the practices being reviewed,
- Clear communication of specific requirements for what constitutes includable effectiveness evidence (i.e., relevant practice and at least one outcome measure/finding, preferably with a baseline comparison),
- Appropriate formal letters of invitation and endorsement of professional, accreditation and industry organizations, and
- Information that meets the needs of relevant IRB chairs and other administrative review offices; assurances of confidentiality when requested.

Organizational Development and Sustainability

- Characterization of the roles and responsibilities of the LMBP Workgroup, Expert Review Panels, and the staff support team evolved over the course of this phase, helping to further specify organizational requirements to support systematic evidence reviews and the production of best practice recommendations on an ongoing basis.

- Several key factors are necessary to support and sustain the development and implementation of the LMBP process:

    - **Transparency**. The process must be open to all relevant stakeholders and the public; no part of it should be conducted behind closed doors. All evidence should be clearly presented and the review process should be clearly defined so that it can be replicated and produce the same results.

    - **Timeliness of recommendations**. Sufficient resources must be allocated to the LMBP process to ensure that reviews are completed in a timely fashion so that recommendations are disseminated while they are still relevant and likely to improve healthcare quality outcomes.

    - **Collaboration**. CDC should not operate independently, but instead should collaborate with existing stakeholder, professional and guideline-setting organizations, as well as those recognized independently as subject matter and methods experts.

    - **Involvement of Partners**. It is critical to ensure that the process be inclusive of not only representation of all laboratory medicine stakeholders but sufficiently responsive to the needs and input of all relevant perspectives and disciplines involved in all phases of the testing process. The partners should be diverse and multi-disciplinary, and must have real opportunities for providing input to impact the LMBP process and outcomes.

    - **Independent Recommending Body**. The evidence review results and identification of evidence-based best practices should be issued by a recommending body that is perceived to be independent, not subject to the influence of any particular faction within the field, the sponsoring agency, nor political considerations.

    - **Organizational Commitment to Sustainability**. The model must be sustainable, with resources available to support the process for the long-term. If the process is perceived as an initiative that will fade away, it will not garner the support necessary to make it effective.

    - **Integration with Existing Efforts (Without Duplication).** A number of organizations are already in the process of identifying and disseminating best practices recommendations. The CDC-led LMBP effort should integrate with these efforts to the extent possible through its evidence-based methods, and should not duplicate them.

RECOMMENDED NEXT STEPS

In moving towards sustained implementation, it is recommended that the Laboratory Medicine Best Practices systematic evidence review and evaluation methods for

assessing the effectiveness of quality improvement practices be further refined and enhanced to include some or all of the following activities.

Methods: *Review Topic Selection*

Refine and standardize the process by which systematic review topics are selected and associated candidate practices are nominated. Topic selection criteria established early in the Initiative's development still apply (burden of problem/quality gap; preventability, availability of existing knowledge, potential effectiveness, operational management, and potential economic benefit), but further refinements are needed in soliciting and responding to suggestions from the field.

Methods: *Analytic Framework*

Refine and standardize methods for schematic representation of a review topic analytic framework for each review question including:

- Formalize a process for establishing functional requirements for practices associated with a selected topic area. A "process mapping" approach may help to outline work flows and common points of intervention at which practices can achieve improvements in healthcare quality outcomes.

- Identify processes from domains of application outside of laboratory medicine that meet the same functional requirements, increasing the likelihood that evidence of effectiveness from these other domains will be regarded as relevant to laboratory medicine practices.

Methods: *Search, Screening and Data Abstraction Methods*

Make further improvements to the review methods and electronic data abstraction tool including:

- Refine, standardize, and document literature search strategy to generate relevant published materials in a broader array of journals and published conference proceedings.
- Develop standardized search and reporting functions for reference and study databases.
- Improve guidance and standardization for screening and abstraction methods for reviewers.
- Refine reviewer/user interface enhancements for data abstraction.
- Structure and formatting of data abstraction template more directly linked with evidence summary templates and individual study evaluation criteria.
- Further standardization of outcome measures, definitions, and their categorization to minimize topic area-specific programming and maximize comparability.
- Develop and implement standardized methods for screening and capturing non-effectiveness evidence related to feasibility of implementation, applicability, economic evaluation and harms and benefits and/or other newly developed criteria.

Methods: *Evidence Summary and Evaluation*

- Finalize evidence summary presentation formats along with development of standardized content and terms to facilitate and ensure consistent evaluations, and when applicable statistical meta-analyses, and recommendation statements (for the LMBP topic area Expert Panels and Workgroup), and for publishing and disseminating evidence reviews and evidence-based recommendations.

- Specify methods for including, evaluating and synthesizing additional non-effectiveness evidence related to implementation feasibility, economic evaluation, applicability (settings, populations, contextual variables) and harms and benefits, incorporating concepts of external validity and internal validity.

- Further refine protocols for nominating, selecting, and guiding the work of expert panelists so that panelists have a clear idea of their roles and responsibilities relative to the Recommending Body and support staff, and panel composition is adequately diversified to represent key stakeholders' perspectives to produce unbiased and scientific evidence reviews.

- Further refine protocols for guiding the work of the LMBP Workgroup (or if not overlapping a Recommending Body) so that members of this body have a clear idea of their roles and responsibilities relative to the expert panelists and support staff.

Network Development for unpublished evidence

- Further develop the network as the principal source for unpublished evidence. Expanding and maintaining this network is essential to the future sustainability of an evidence-based laboratory medicine practice recommendations process, as the main challenge to its success remains insufficient published evidence.

- Further refine guidance to network participants on informational requirements for submitting evidence.

- Develop and implement an education / curriculum strategy that familiarizes laboratory managers with methods for improving the quality of unpublished process improvement / quality assurance studies so that data from these studies are consistently available to inform "best practice" recommendations.

- Expand strategies to extend the breadth and depth of the network to provide greater opportunities for identifying participating organizations and individuals within those organizations responsible for relevant practice evaluations and quality improvement initiatives.

- Maintain a network tracking database with strategic information to facilitate contacts, targeted follow-up as well as routine communication with network affiliates.

Organizational Development and Sustainability

- Create a specific business plan for implementation and funding alternative models based on collaboration with key stakeholders.

- Develop and implement communication, publication and other dissemination strategies based on collaboration with key stakeholders to optimize impact of evidence reviews and further the implementation of evidence-based methods and standards for quality improvement in laboratory medicine.

Development of a process for assuring a pipeline of future topic areas and priorities for evidence reviews based on broad stakeholder engagement, including identification of appropriate evidence.

ACKNOWLEDGMENTS

The following report represents the collective efforts of many people dedicated to improving the quality of laboratory medicine. The authors would like to recognize a number of people whose support, guidance, expertise, and commitment have made it possible to carry out the task to developing and implementing an evidence-based process for identifying best practices.

At the Centers for Disease Control and Prevention and the Department of Health and Human Services, many people have been committed to the development of this process. This includes the thoughtful project oversight and guidance by Drs. Roberta Carey, Barbara Zehnbauer, Julie Taylor, and Shambavi Subbarao, and their predecessors in the Division of Laboratory Systems, Joe Boone and Devery Howerton.

We are extremely grateful for the time and dedication to this process extended by our Workgroup members. Their careful consideration of each step in the process, coupled with their commitment to improving laboratory medicine was fundamental in completing the pilot phase. We extend our appreciation to Raj Behal, MD, MPH; John Fontanesi, PhD; Julie Gayken, MT(ASCP); Cyril ("Kim") Hetsko, MD, FACP; Lee Hilborne, MD, MPH; James Nichols, PhD; Mary Nix, MS, MT(ASCP)SBB*;* Stephen Raab, MD; Ann Vannier, MD; and Ann Watt MBA. We also appreciate the contributions of our two *ex-officio* members Sousan S. Altaie, PhD of the U.S. Food and Drug Administration and James A. Cometa of the Centers for Medicare and Medicaid Services**.**

In addition to the Workgroup members, we would like to acknowledge the evidence review and methodological guidance provided by expert panelists Steve Kahn, PhD; Paul Valenstein, MD, FCAP; Denise Geiger, PhD; David Hopkins, MD, MPH; Ronald Schifman, MD, MPH; Corinne Fantz, PhD; Dana Grzybicki, MD, PhD; Kent Lewandrowski, MD, PhD; Rick Panning, CLS(NCA), MBA; Dennis Ernst, PhD; Margret Oethinger, MD, PhD; and Melvin Weinstein, MD.

A number of laboratories agreed to consider our request to make available unpublished evidence for the pilot test of our systematic review methods, including Bay State Health Systems, Colorado University's Cancer Care Center, Emory Healthcare, Geisinger Health Systems, Johns Hopkins Medical Center, LBJ Hospital, Loyola University Medical Center, Mather Hospital, Memorial Health Systems, Providence Health Care, Regions Health Care, SonoraQuest, the Southern Arizona Regional Veterans Affairs Medical Center, the University of Kansas Medical Center, the University of Maryland Medical Center, and the University of Washington Medical Center.

Library services were provided by Ms. Janette Schueller, MLS.

CONTENTS

EXHIBITS

FIGURES

TABLES

1.0 PROJECT OVERVIEW

1.1 PURPOSE AND BACKGROUND

Clinical laboratory services play a vital role in the delivery of individual health care and public health in the United States. The Department of Health and Human Services' (HHS) Centers for Medicare and Medicaid Services (CMS) certifies over 200,000 U.S. laboratories under the provisions of the Clinical Laboratory Improvement Amendments of 1988 (CLIA).[4] These laboratories provide more than 1,000 laboratory tests for human conditions, and about 500 of these tests are used daily.

In response to the Institute of Medicine's call to improve quality in medicine (Institute of Medicine 2000, 2001), CDC's Division of Laboratory Science and Standards (DLSS) in the Office of Surveillance, Epidemiology, and Laboratory Science (OSELS) is supporting the development of a systematic, evidence-based process, based on transparent methods to identify best practices in laboratory medicine. This initiative targets the pre- and post-analytical phases of the laboratory total testing process (Barr and Silver 1994), as these phases encompass the majority of laboratory-related errors and opportunities for improvement. This effort began in October 2006, when CDC convened the Laboratory Medicine Best Practices Workgroup (LMBP Workgroup), a multidisciplinary advisory panel comprising experts in several fields of laboratory medicine, clinical medicine, health services research, and health care performance measurement. The LMBP Workgroup was supported by a team from DLSS and its contractor, the Battelle Centers for Public Health Research and Evaluation (Battelle). The overall goal of the effort is to develop methods for completing systematic evidence reviews and evaluations for making evidence-based "best practice" recommendations for practices with demonstrated effectiveness to improve the quality of health care and patient outcomes. These evidence reviews and recommendations will assist professional organizations, government agencies, laboratory professionals, clinicians, and others, who provide, use, regulate, or pay for laboratory services to make decisions to improve health care quality based on evidence of effectiveness.

To date, the LMBP methods development process has completed three phases. Phase 1 (October 2006-September 2007) involved a "proof of concept" test of an approach to searching, screening, and evaluating evidence as the basis for best practice recommendations. An outcome of Phase 1 was to act on a Workgroup recommendation and enlarge the search for evidence to unpublished assessments performed for the purposes of quality assurance, process improvement and/or accreditation documentation, and to adapt conventional systematic review methods to allow inclusion of unpublished quality improvement studies. Phase 2 (September 2007-November 2008) involved a pilot test of further refined methods to obtain, review, and evaluate published and unpublished evidence, along with collecting observations via key informant interviews about organizational and implementation issues successfully addressed by other recommending bodies about the development and dissemination of guidelines and recommendations. Phase 3 (September 2008-February 2010) used feedback and results obtained in Phase 2 to refine the data collection instruments and study rating methodology to better address the material available in laboratory medicine studies,

[4] Centers for Medicare and Medicaid Services (http://www.cms.hhs.gov/clia/) [accessed February 1, 2010]

including meta-analysis of practice effect size. In addition, a standard cycle in evidence-based medical practice reviews (Ask, Acquire, Appraise, Analyze, Apply, and Assess) was adapted by the project to introduce the "A-6 Cycle," to include an analysis step (Shaneyfelt et al. 2006).

1.2 PHASE 3 OBJECTIVES

More specifically, the project's Phase 3 had three objectives:

- Refine, further develop and pilot test methods that had been evaluated initially in the proof of concept and initial pilot test phases.
- Test the feasibility of developing a national network of facilities that would agree to furnish unpublished studies for use in quality improvement, practice-specific systematic evidence reviews.
- Recommend an approach to implementing on a sustainable basis the process for systematic evidence reviews and identification of best practices for laboratory medicine, including developing a network of organizations to provide unpublished evidence.

1.3 LABORATORY MEDICINE BEST PRACTICES WORKGROUP

Continuing their work from the initial Proof-of-Concept phase, the LMBP Workgroup consists of 13 invited members, including two ex officio representatives from the Centers for Medicare and Medicaid Services (CMS) and the Food and Drug Administration (FDA). The Workgroup members are clinicians, pathologists, laboratorians, and specialists in systematic evidence reviews with recognized expertise in performance measurement, standard setting, and health services research. The Workgroup members' main functions are to provide overall guidance and feedback on developing review and evaluation methods for making evidence-based best practice recommendations. As the "Recommending Body" for the LMBP pilot test, the Workgroup reviewed, provided guidance and made recommendations on:

- topic area selection and criteria for practice reviews (ask)
- recruitment of Laboratory Medicine Best Practices Network affiliates for unpublished studies (acquire)
- format and content for evidence summaries and draft corresponding best practice recommendations (analyze) prepared by Expert Panels and CDC /Battelle Review Team staff
- evaluation methods (analyze) for producing evidence-based best practice recommendations
- strategies and methods for presenting and disseminating recommendations (apply)
- systematic evidence review methods used by Expert Panels and the CDC/Battelle Review Team to acquire, appraise and analyze published and unpublished studies

- strategies and alternatives for implementing an organizational structure for routine and sustainable use of the Laboratory Medicine Best Practices methods to produce systematic evidence reviews of laboratory medicine quality improvement practices

1.4 ORGANIZATION OF THIS REPORT

The following sections summarize work completed during Phase 3 and the LMBP methods using the "A-6" cycle steps. Section 2 describes the selection of review topics, including selection criteria applied and the topics chosen. Section 3 outlines the systematic review methods developed and employed during Phase 2 and the pilot test, including the development of an analytic framework and one or more focused review questions(ASK); the search strategy for evidence from the published literature and unpublished sources (ACQUIRE); the screening of acquired studies then abstraction and standardization of information from individual studies (APPRAISE); the analysis and rating of an aggregated body of evidence (ANALYZE); and the translation of evidence-based findings and best practice recommendations into practice (APPLY). Section 4 presents the pilot test results of the evidence reviews for practices associated with three topics, "Patient Specimen Identification," "Critical Values Test Result Reporting and Communication," and "Blood Culture Contamination." Section 5 reports Phase 3 findings about the need for further refinements in evidence collection, review and evaluation methods, enhancements needed in network development and outreach, and strategic goals for organizational development and implementation planning. A set of appendices is included; (A) lists the 2009 Workgroup members, (B) lists the three Evidence Review Panel members, (C) describes the roles and responsibilities of the Workgroup and Review panels, (D) details the literature search strategies used, (E) presents the detailed evidence review summaries and quality ratings, (F) provides the guidance given to panelists for rating study quality, and (G) the guidance given to panelists for rating effect size, and (H) the record structure and coding guidance for the data abstraction database.

2.0 PHASE 3 TOPIC SELECTION

For the purposes of the pilot phase, three topic areas were selected, based on the following criteria. To be selected, a topic area was required to:

- address a defined quality issue/problem in laboratory medicine consistent with the six IOM healthcare quality aims (safety, timeliness, effectiveness, equity, efficiency, patient-centered),
- be framed by at least one focused review question,
- be associated with at least three potential practices that attempt to improve performance/quality outcomes related to the defined quality issue/problem,
- have outcome measures of broad stakeholder interest that can be used to assess practice effectiveness, and
- have evidence (studies/data) of practice effectiveness available from published sources and potentially from unpublished sources.

In consultation with the Workgroup, a decision was made to continue with the topic areas previously selected for use in the earlier Proof-of-Concept (Phase 1) and initial pilot test (Phase 2); (Patient Specimen Identification, and Communication of Critical Value Test Results), and to add a topic area (Blood Culture Contamination) that also met the selection criteria.

2.1 TOPIC 1: PATIENT SPECIMEN IDENTIFICATION

Quality Issue / Problem: Patient specimen identification errors may contribute to adverse patient events and wasted resources.

Review Question: What are effective interventions/practices for reducing patient specimen and/or test result identification errors?

Potential Interventions / Practices: Earlier reviews of published and unpublished evidence indicated that sufficient evidence would likely be available to consider the effectiveness of one practice in two care settings:

- **Barcoding Systems** - Electronic bar-coding on both patient and specimen used to establish positive identification of specimen as belonging to patient.
- **Point-of-Care-Testing Barcoding Systems** - Automated patient and sample/test identification system when diagnostic testing is conducted using a testing device at or close to the patient.

Possible Outcome Measures:

- Specimen and/or test result identification errors (rates), and
- Repeat testing (rates) due to ambiguous patient specimen/test result identification.

2.2 TOPIC 2: COMMUNICATION OF CRITICAL VALUE LABORATORY TEST RESULTS

Quality Issue/problem: The reporting of critical/panic value laboratory test results that are incorrect, incomplete, and / or untimely can result in ineffective communication, which may contribute to patient adverse events.

Review Question: What practices are effective for timely and accurate communication of laboratory critical test results to responsible / licensed caregivers?

Potential Interventions / Practices: Earlier reviews of published and unpublished evidence indicated that sufficient evidence would likely be available to consider the effectiveness of two practices:

- Automated notification of critical value test results via computerized alerting systems and/or personal electronic devices (e.g., alphanumeric pagers or SMS ‘text’ messaging), and
- Customer Service (or “Call”) center.

Possible Outcome Measures:

- Time to receipt: Documented time from laboratory confirmation of test result to caregiver receipt of result,
- Time to treatment: Length of time from laboratory confirmation of critical result to resolution by clinical staff, and/or
- Accuracy/error rate in confirmation of telephone-reported results.

2.3 TOPIC 3: BLOOD CULTURE CONTAMINATION

Quality Issue/problem: Blood culture contamination may lead to false positive cultures that, in turn, lead to inappropriate follow-up and treatment

Review Question: What practices are effective for reducing blood culture contamination?

Potential Interventions / Practices: Initial reviews of published evidence indicated that sufficient evidence would likely be available to consider the effectiveness of three practices:

- Dedicated Phlebotomy Teams: Staff certified draw blood for laboratory tests.
- Pre-packaged Prep Kits: Pre-packaged aseptic supplies that are prepared in-house or commercially purchased.
- Venipuncture (vs. Intravenous Catheter): Puncture of a vein through the skin to withdraw blood (*vs.* use of a thin flexible tube inserted into the body).

Possible Outcome Measures:

- Blood culture contamination rate – number and proportion of blood cultures growing contaminant organisms, and or
- Positive Predictive Value (less direct outcome measure).

3.0 SYSTEMATIC REVIEW METHODS

This section summarizes the methods developed and piloted to collect, screen, review and evaluate evidence from published and unpublished sources. In Phase 3, the A5 evidence-based laboratory medicine cycle (see, e.g., Price, Glenn, & Christenson, 2009) was adapted by including a sixth step (Analyze), to describe the review process used to identify best practices for laboratory medicine. The CDC-LMBP "A6 cycle" steps are:

(1) ASK a focused question(s) in the form of a quality issue problem statement;
(2) ACQUIRE evidence by identifying sources and collecting potentially relevant studies;
(3) APPRAISE studies by applying screening criteria then abstracting, standardizing and rating information from included studies;
(4) ANALYZE by rating the evidence base using meta-analytic techniques when feasible

a. Expert panels use the evidence summaries provided in Evidence Summary Tables and standardized findings to reach consensus on the study quality and effect size magnitude ratings to transparently translate the findings for each practice into a draft evidence-based recommendation;
b. These evidence reviews become the basis for the practice recommendations reached by the Laboratory Medicine Best Practices Workgroup (serving in its capacity as the "Recommending Body")

(5) APPLY by disseminating evidence review findings and recommendations via peer-reviewed literature and other media, educational programs, and guidelines as appropriate, to influence and facilitate actual practice implementation to improve quality;

(6) ASSESS practices to evaluate implementation performance outcomes/results to evaluate whether and to what extent quality improvement occurred, determine the applicability of practices to various settings or other important implementation characteristics, and consistent with continuous quality improvement, identify other quality issues that can be framed as new opportunities for asking questions that can be addressed by either new reviews and/or updated reviews to continue the cycle of improvement.

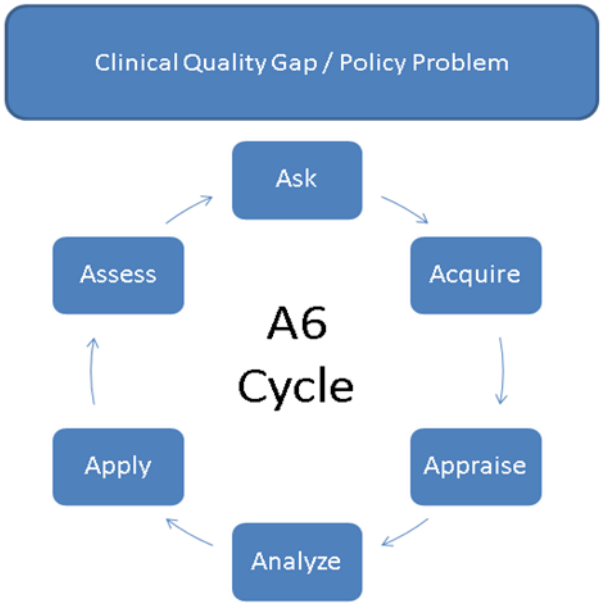


FIGURE 1. THE EVIDENCE-BASED PRACTICE CYCLE ADAPTED FOR LABORATORY MEDICINE

This general sequence of LMBP systematic review activities leading to recommendations is described in Figure 2 and essentially follows the sequence outlined

by Khan, ter Riet, Glanville et al. (2001) and those used by the Community Guide (Zaza, Briss, and Harris 2005) and US Preventive Services Task Force.[5]

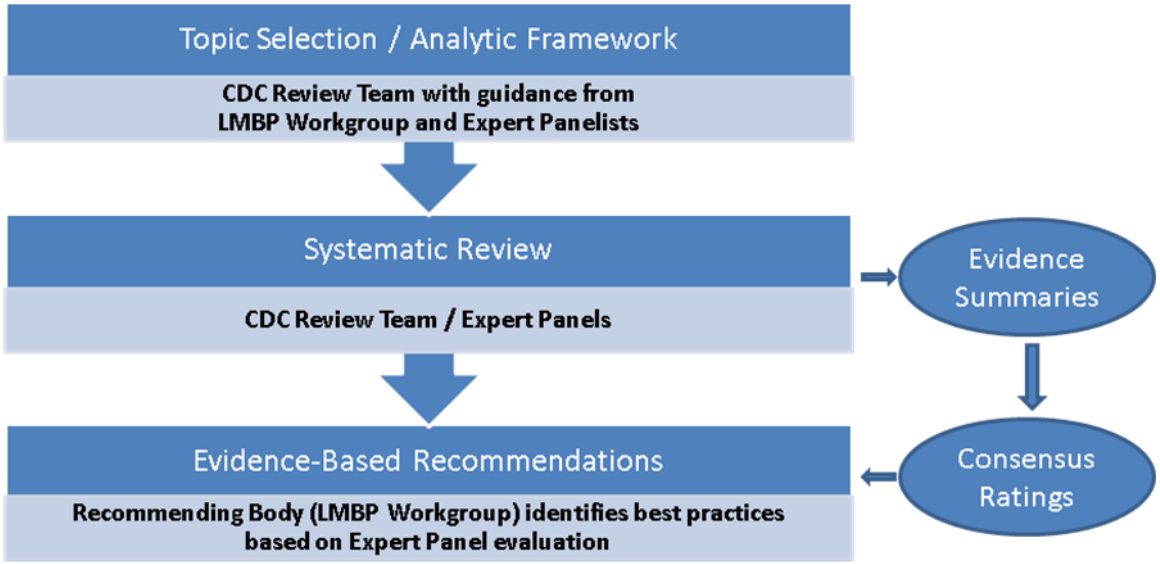


FIGURE 2. GENERAL SEQUENCE FOR FORMULATING EVIDENCE-BASED BEST PRACTICE RECOMMENDATIONS

3.1 STEP 1 – ASK: DEVELOPING AN ANALYTIC FRAMEWORK

The initial step of an evidence review once a topic area has been screened and selected is to ASK one or more focused questions in the form of a problem statement, and develop an analytic framework to clarify and define the scope of the review. The generic framework consists of a set of basic elements that correspond to the criteria used in selecting topics and relevant evidence for review. Completing an analytic framework is consistent with the Institute of Medicine definition of the quality of care (the degree to which health care services for individuals or populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge) by characterizing a laboratory medicine topic as it relates to a quality issue in need of improvement. The analytic framework facilitates framing systematic review questions that can be addressed by evidence by specifying these elements:

- Quality Issue / Problem that can be framed by:
  - Evidence of a defined quality gap that can be improved or prevented
  - Review Question (linking quality issue/gap, interventions/practices, and outcome measures);
- Potential Interventions / Practices that may improve quality

[5] For the most up-to-date overview of methods used by the US Preventive Services Task Force, consult the Agency for Healthcare Research and Quality (http://www.ahrq.gov/clinic/uspstmeth.htm) [accessed February 1, 2010].

- Outcome Measures (intermediate and health-related outcomes) of interest
- Additional Harms and Benefits associated with implementing the intervention/practice

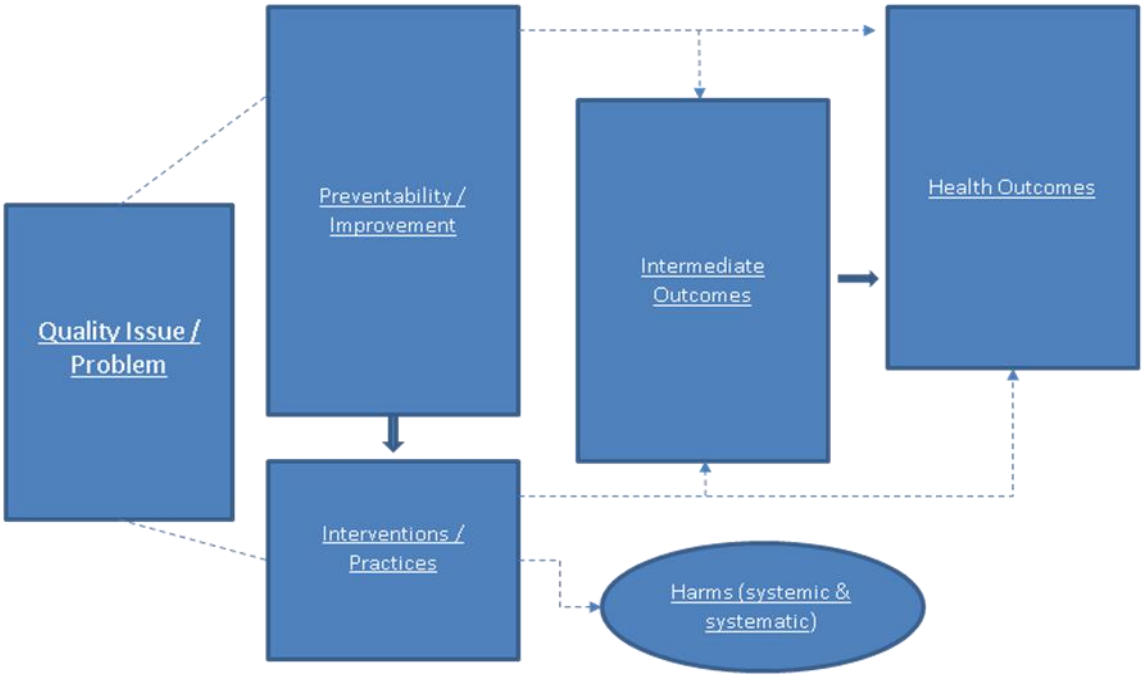


FIGURE 3. LABORATORY MEDICINE BEST PRACTICES – BASIC ANALYTIC FRAMEWORK

An initial analytic framework is based on a preliminary review of published literature, and is refined using additional information obtained as the evidence review progresses. Figures 4a, 4b, and 4c depict the analytic frameworks used to guide the three systematic reviews.

FIGURE 4A. ANALYTIC FRAMEWORK FOR PATIENT SPECIMEN IDENTIFICATION

Review Question: What are effective interventions/practices for reducing patient specimen identification errors?

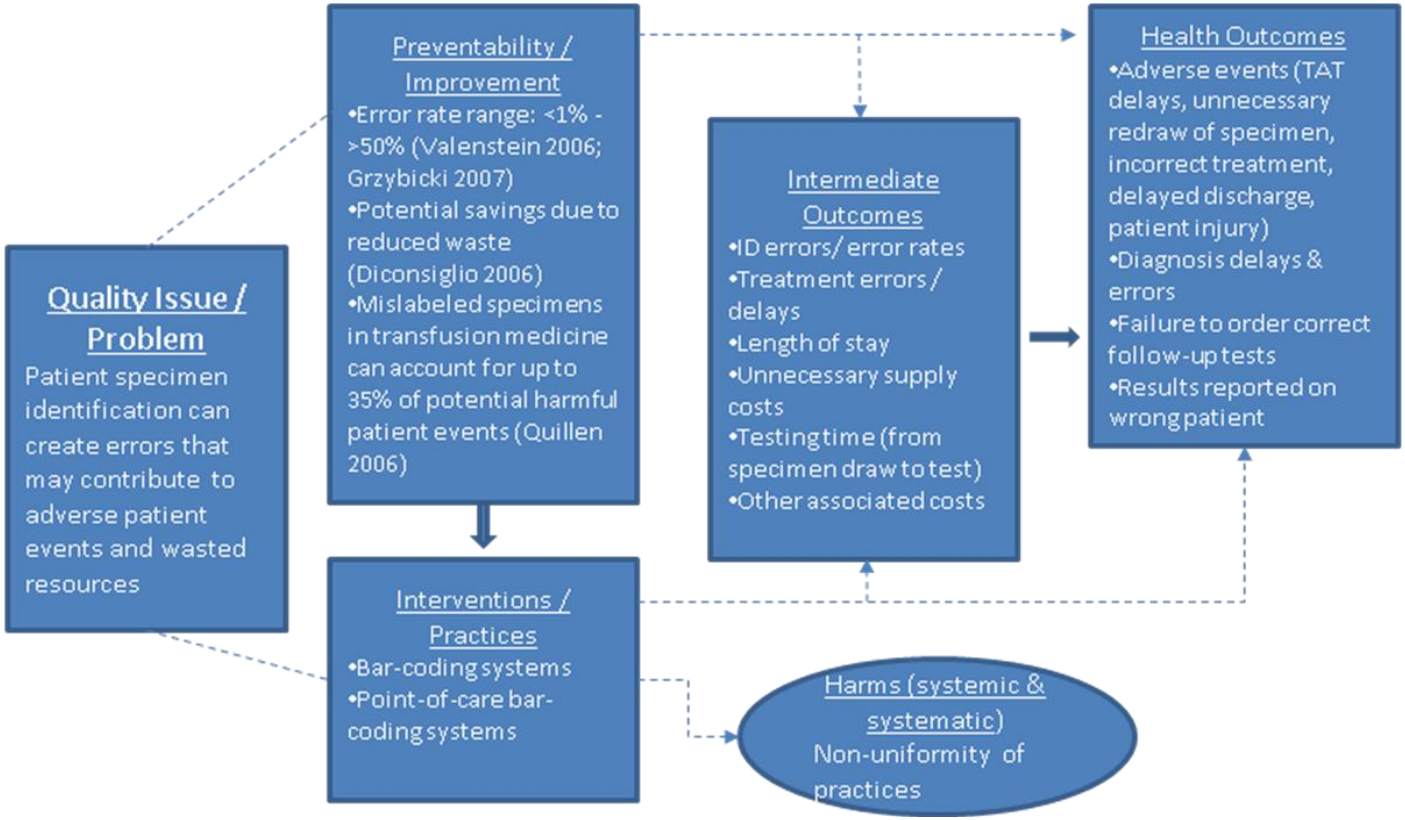


FIGURE 4B. ANALYTIC FRAMEWORK FOR CRITICAL VALUES REPORTING & COMMUNICATION

Review Question: What practices are effective for timely and accurate communication of laboratory critical test results to responsible/ licensed caregivers?

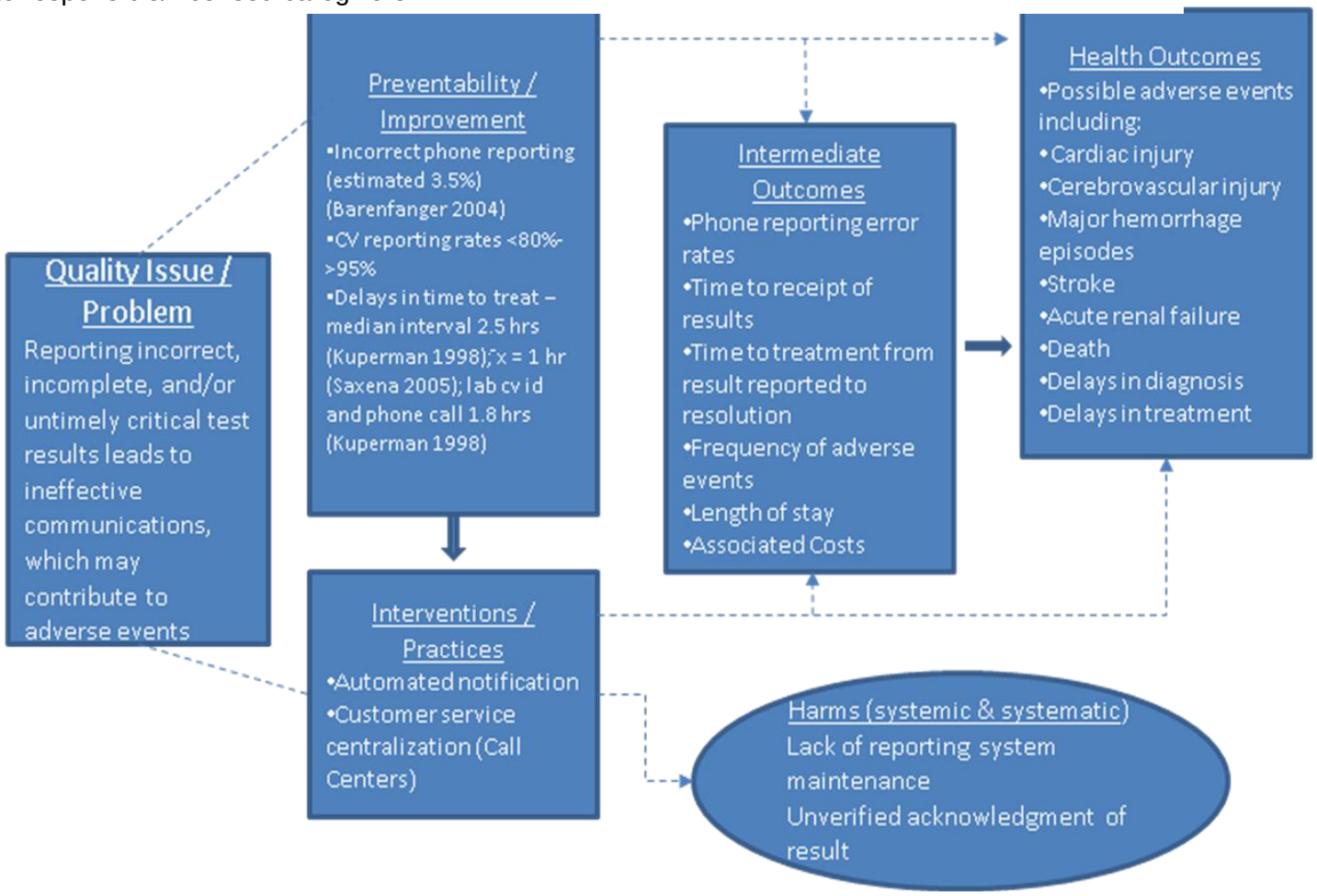
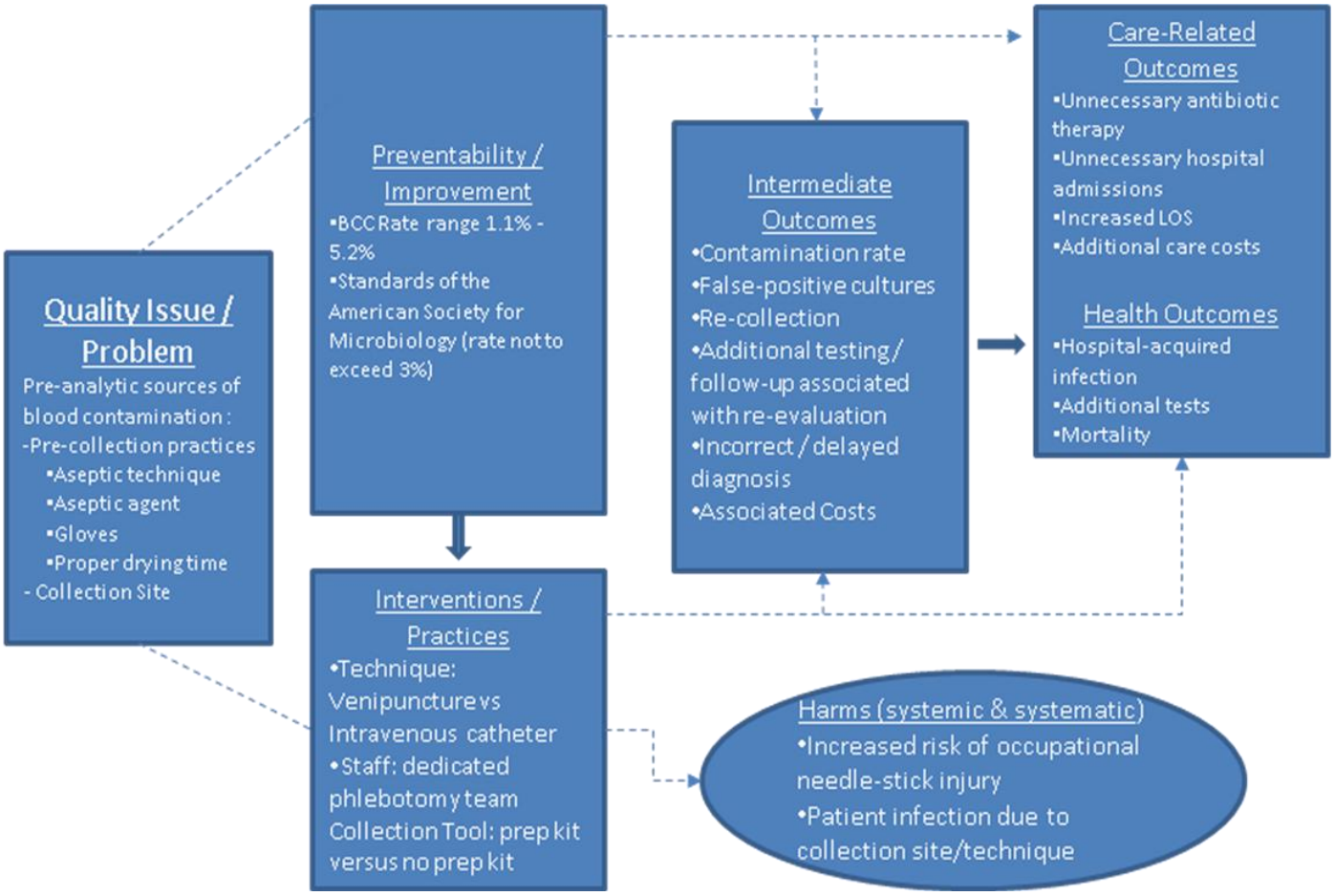


FIGURE 4C. ANALYTIC FRAMEWORK FOR BLOOD CULTURE CONTAMINATION

Review Question: What practices are effective for reducing blood culture contamination?



3.2 STEP 2 – ACQUIRE THE EVIDENCE

3.2.1 PUBLISHED LITERATURE

Consistent with established systematic review methods, ACQUIRING the evidence requires developing a search protocol applied to electronic databases, as well as other means and sources including hand searching of bibliographies, correspondence with experts in the field to identify published studies that assess candidate practices for each review topic. In each case, the review question(s) and analytic framework established in the ASK (A-1) step guided the selection of initial search terms (see Appendix D for details of the Phase 3 pilot test literature search strategies).

Conducted with the assistance of a professional librarian, for all three topics the search strategy involved a comprehensive search of English-language literature published during or after 1996 using multiple databases and other strategies. These included:

- PubMed, MedLine, CINAHL, BMJ Clinical Evidence, and Cochrane databases,
- Professional guidelines electronic databases (AHRQ, Cumitech, CLSI, ISO, NACB),
- Hand searching journals of relevance to the review topic, reports, conference proceedings, and technical reports,
- Reference lists of relevant published studies, reviews, and other sources (e.g., reports, presentations, guidelines, standards), and
- Key informants: consultation with Expert Panel and Workgroup members for relevant information sources.

3.2.2 UNPUBLISHED EVIDENCE SEARCH

One of the principal findings of the project's Proof-of-Concept (Phase 1) was that considerably more evidence might be available outside of the published and peer-reviewed literature. It was observed that practices in laboratory medicine are not often subjected to experimental trials, controlled, or observational studies to assess their effectiveness before they are implemented. Such formal studies are hard to do, expensive, commonly impractical and thus difficult to justify. However, laboratories, hospitals, and other health care institutions often conduct less formal analyses and assessments of information that they collect routinely before and after they adopt new practices or change established practices, especially if the proposed changes involve reorganizing the way the laboratory works, changing management systems, or adding new resources (systems, instruments, people). Typically, these assessments are not called "studies" or "research", but they may be rigorous and objective evaluations of high-quality data and thus constitute evidence of practice effectiveness. A key Phase 2 objective was to develop and implement methods for incorporating these unpublished practice assessments as studies in the systematic evidence reviews. As such, unpublished studies are reviewed and evaluated according to the same criteria and standards as published evidence.

Search methods implemented for unpublished evidence included the following steps:

1. Obtained the support and endorsement of key stakeholder organizations to encourage clinical laboratories and healthcare organizations to participate in the pilot test. During Phase 3, endorsements were obtained from and presentations or materials soliciting participation were made available to the following organizations, their newsletters, and at the following meetings:
   a. Clinical Laboratory Management Association's ThinkLab
   b. American Society for Microbiology
   c. American Association for Clinical Chemistry
   d. American Society for Clinical Laboratory Science
   e. Clinical Laboratory Improvement Advisory Committee

2. Identified facilities likely to have completed relevant assessments, based on:
   a. Conference papers or other public presentations
   b. Relevant publications that implied the author might have additional data beyond what was reported (e.g., more recent data, or data more encompassing in scope or care setting)
   c. Personal knowledge of our Workgroup members.

3. For those facilities which had likely completed relevant assessments, identified and contacted a senior laboratory scientist, laboratory director, or other appropriate representatives (e.g., involved in patient safety, quality management, clinical research, regulatory/accreditation compliance) to describe the aims of the project and explore the circumstances under which the organization would consider participating in the pilot test.

4. Provided additional information about the pilot test to the facility point-of-contact to share with colleagues and obtain a preliminary assessment from the organization's Institutional Review Board (IRB) chair for release of previously completed studies with de-identified data.

5. Extended a formal invitation to the organization, providing more general guidance about the type of information needed for unpublished studies.

6. Established any formal confidentiality safeguards or conditions under which the information would be provided for the purposes of the pilot test of systematic review methods.

7. Reviewed study information and other material received, and follow-up with additional information requests as needed.

To minimize the burden on pilot test participants and maintain the consistency with published evidence, only previously completed studies were requested (i.e., no new data), and it was suggested that these studies may be derived from multiple sources, including internal assessments, case studies, Failure Mode and Effects Analyses (FMEA), and quality improvement studies. Facilities were also requested to provide data that contained no personal health information concerning patients. A commitment was made to de-identifying all data and studies submitted, and each facility offered the option to remain anonymous in the summaries describing pilot test findings. All organizations that requested anonymity when providing unpublished studies remained anonymous in

the final evidence summaries (Appendix E) used by the Expert Panels and the Workgroup.

Using this approach in Phase 3, initial exploratory discussions were held with representatives from 37 facilities (Step 3). Following these initial discussions, formal invitations were issued to 9 organizations (Step 5) to provide studies for each of the three topic areas (27 invitations in total), and 23 submissions were received. Ultimately, after subjecting the submissions to the same exclusion and inclusion criteria applied to published literature as detailed in the previous section, this approach resulted in about half (12) of the unpublished studies being included in the systematic review evidence base for the three topic areas (Patient Specimen Identification: 6; Critical Value Test Result Reporting: 4; Blood Culture Contamination: 2)

3.3 STEP 3: APPRAISE – SCREEN, ABSTRACT AND STANDARDIZE

LMBP review methodology includes the screening of all information obtained in the ACQUIRE step by two independent reviewers.

Two reviewers independently screened information acquired from literature searches and from submitted unpublished studies by applying inclusion and exclusion criteria as detailed below. A pre-abstraction reference list of literature meeting the initial inclusion criteria was generated, indicating references that would be considered for full-text review.

3.3.1 EXCLUSION CRITERIA

Upon review of the title and abstract of an article or an unpublished submission, it was excluded if one or more of the following exclusion criteria were applicable.

- No practice was assessed (i.e., no outcome measures were identified)
- The practice was not sufficiently described
- The content was a commentary or opinion piece

3.3.2 INCLUSION CRITERIA

An article or unpublished submission was included for a full-text review if at least one practice was described that appeared to satisfy all of the following inclusion criteria.

- Relevant to the review question
- Satisfied practice-specific criteria (characteristics and requirements)
- In use and available for adoption
- Reproducible in other comparable settings
- Addresses a defined/definable group of patients
- Has a potential impact on an outcome related to at least one of the following IOM healthcare quality aims: effectiveness, efficiency, patient-centeredness, safety, timeliness or equity

Figures 5 a-c provides a summary of search and screening results for the LMBP Phase 3 pilot test three topic areas. The list of 598 references included in the initial screening for Patient Specimen Identification, Figure 5a, resulted in a total of 16 articles that met

the inclusion criteria for use in the systematic review and ultimately 9 that were included in the body of evidence. The list of 540 published references included in the initial screening for Communicating Critical Values, Figure 5b, ultimately resulted in a total of 5 articles included in the body of evidence. 1677 published references concerning blood culture contamination ultimately yielded 14 articles that could be used (Figure 5c). While this rate of reduction may seem restrictive, it is quite consistent with rates observed in other systematic reviews (Horvath and Pewsner 2004:25-26).

All studies meeting the screening criteria are then subject to full-text appraisal by abstracting and standardizing study information to prepare evidence summaries. This compilation of individual studies related to a practice generates a body of evidence that is used by review staff and expert panelists to complete the ANALYSIS step. For each study, this process consists of (1) data abstractions to standardize study information, independently conducted by at least two reviewers; (2) a reconciliation and consensus of data abstractions where there was not complete agreement; (3) when appropriate, calculation of a standardized effect size for each individual study's observed effects (typically using either an Odds Ratio or Cohen's-d statistic, depending on the nature of the data), and (4) summarization and synthesis of the practice body of evidence in a standardized evidence summary table for used by the expert panelists to complete the practice evidence reviews and evaluations. Once each study was abstracted and the evidence rated, a summary Body of Evidence Table and graphic representation using forest plots of study results for each practice was created that summarized an overall summary effect across studies and overall consistency of studies included in the body of evidence (See Appendix E).

FIGURE 5A-C. SEARCH RESULTS FOR PHASE 3 TOPIC AREAS

Figure 5a. Topic Area: Patient Specimen Identification Literature Search Results

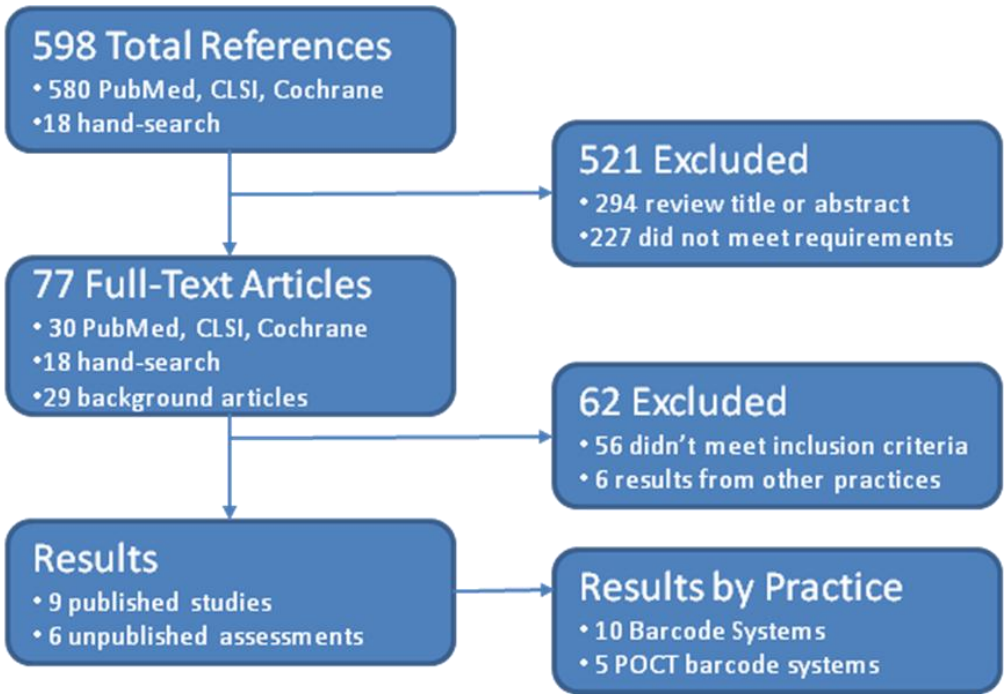


Figure 5b. Topic Area: Critical Value Reporting Literature Search Results

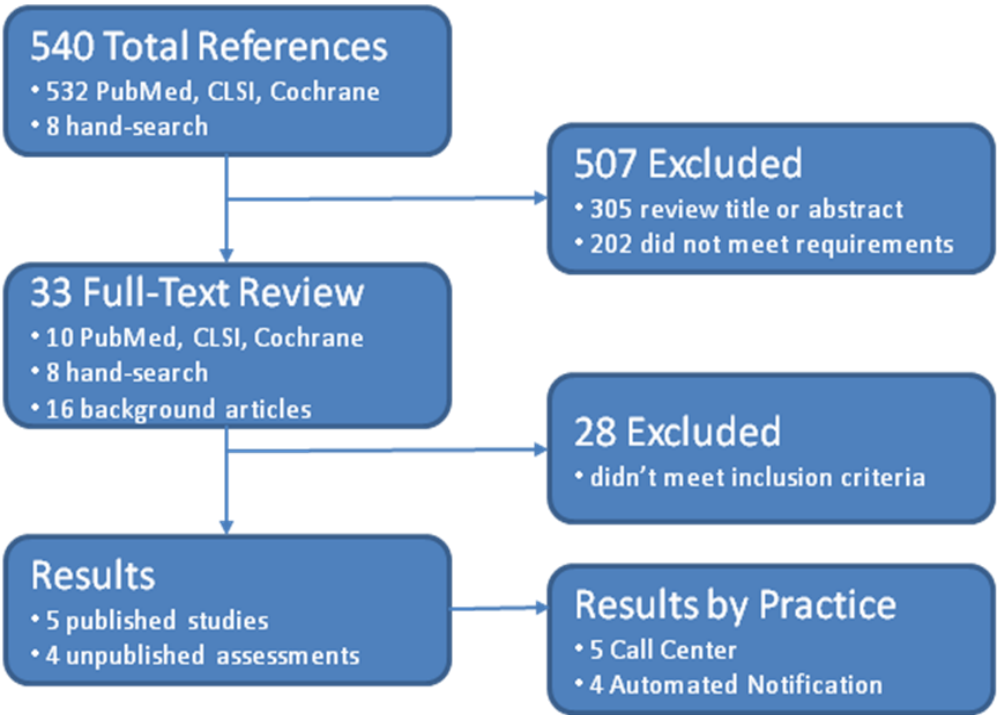
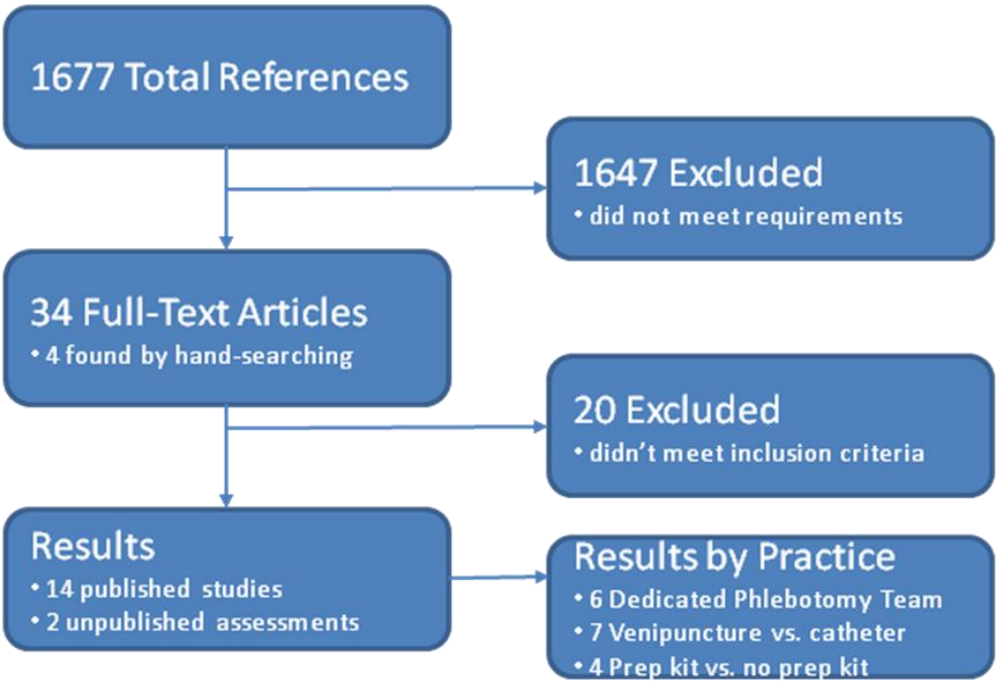


Figure 5c. Topic Area: Blood Culture Contamination Literature Search Results



3.3.3 DATA ABSTRACTION AND EVIDENCE SUMMARY

Published and unpublished studies are not reported in a uniform format, making it necessary to consistently abstract from each one the relevant information in a standardized form for the data elements required for evaluation of study quality and effect size. A primary goal of the data abstraction tool is to guide the systematic rating of study results so that standardized information is used to develop consistent, transparent and well-supported ratings for each study. To avoid biases, two reviewers are assigned to complete this abstraction task independently using a standardized abstraction form, and then compare their results. If any divergence appears, the reviewers discuss their rationale and arrive at a consensus result, at times with assistance from additional independent reviewers. Typically, such differences are due to ambiguous reporting on the part of the study authors, or that the study was completed to satisfy some objective other than answering the review question.

In Phase 2, an electronic standardized data abstraction tool was developed to produce standardized data abstractions of the information required to make judgments of the four dimensions of study quality and evaluate effect size. In Phase 3, this data abstraction tool was refined to make the abstraction process more consistent across reviewers, the information abstracted more standardized and efficient and with respect to completion of the evidence summary tables and application of the study quality rating criteria. Detailed information on the data abstraction tool is provided in Appendix H – Data Abstraction Codebook. These improvements resulted in greater consistency of data abstraction results across studies, and a more transparent process of rating study quality and findings.

The abstraction tool consists of five parts, one providing bibliographic information, and the others for assessing dimensions of study quality. These dimensions, and their component measures, were adapted from existing study quality rating instruments and theory to best capture the study and reporting conventions in typical laboratory medicine quality improvement studies. As such, they focus less on the internal validity of the study with respect to causal inference, and put greater weight on the accuracy of the evidence obtained from the methods and measures, assessment of sources and potential for bias from sources outside the practice being tested, and documentation of the generalizability of quality improvement study results. The items and guidance for recording and rating study quality data are reported in Appendix F. The main parts include:

- *Bibliographic information* for published studies and other source information for unpublished studies
- *Study characteristics* (design, sample, time period, care setting) that may be important for contextualizing the results, identifying study quality limitations, and for assessing the practice's applicability to a wide range of care settings
- *Practice characteristics,* including what may be important for assessing the adequacy of practice description with respect to content, implementation, population / practice setting, staff, training, resource, process and functional requirements, and costs associated with implementing the practice.

- *Outcome measure characteristics*, that capture the accuracy and completeness of the evidence collected to estimate the impact of a practice on one or more outcomes. As studies often report more than one outcome associated with implementing a practice, in Phase 3 the convention of using statistical meta-analysis to evaluate only the outcome(s) that most directly address the review question related to the IOM domains of healthcare quality (i.e., safe, timely, effective, patient-centered, efficient, and equitable) was employed.

- *Results*, including findings for all applicable outcome measures reported, including both practice effectiveness/quality outcomes associated with IOM domains and findings related to practice applicability, cost, feasibility and implementation issues, and other harms and benefits.

Once the data from each study were abstracted in detail, a less detailed Evidence Summary Table and draft ratings of the quality of evidence in each study part (along with a justification for the rating if points were deducted) for each was prepared to facilitate communication of study quality ratings. These Evidence summary tables are presented in Appendix E.

3.3.4 STANDARDIZING THE EFFECT SIZE

Little if any of the evidence available for the included practices was based on randomized designs. The typical LMBP study uses a pre-post one-group design. That is, the study provides an estimate for the outcome that resulted from a previous standard or "comparison" practice and an after-implementation estimate of the new or "tested" practice on the same measure. Typical outcome measures include the practice error rate, the proportion meeting a timeliness threshold, receipt of appropriate care, time to acknowledge critical information.

In contrast with controlled research, the comparison practice against which a new practice is tested likely varies across studies. This can affect the difference score (the finding) obtained as much as the new practice. When interpreting magnitude of effect, consideration is given to the actual practices being compared as well as the potential that other sources of influence (e.g., implementation, changes in practice setting, staffing, training, etc.) may distort the difference observed in a finding. In comparative effectiveness research, the findings represent the difference between practices as implemented in an uncontrolled natural setting. If there are great differences in the comparator practices contributing to an evidence summary the results obtained from the trial may not be representative of the impact of the new practice over a common base. This typically presents as a lack of consistency in findings given a common new practice.

To facilitate comparability in evaluating diverse outcome measures and practice comparators, and aid reviewers in judging the magnitude of effect between a new/tested practice and a comparison practice, study results were transformed to a common metric (know generally as an 'effect size'). When outcome measurement represented a dichotomous outcome (e.g., presence or absence of a blood culture contaminant), odds ratios (or occasionally logged odds ratios) were calculated. When results from

continuous measures were being recorded (such as time to an event), Cohen's *d* was adopted to represent the findings.[6]

(1) Odds Ratio (OR) compares the chance of an event occurring in one group versus another group (e.g., new/post-practice versus standard/pre-practice) for dichotomous outcomes (i.e., 2 possible outcomes such as yes/no; error/no error) and has the following interpretation:

- OR > 1: new practice is more successful than the standard practice; the larger the number, the greater the relative success
- OR = 1: new practice is equal to the standard practice,
- OR<1: new practice is less successful than the standard practice; the smaller the number, the worse its relative success

For example, OR = 1.5 means the tested practice is half again as successful as the old practice, OR = 2.0 means the tested practice is twice as successful as the standard practice, and OR 0.66 means the tested practice is two thirds as successful as the standard practice. The logged odds ratio is simply the log of the odds ratio and centers on zero instead of 1 (i.e., 0 = no difference, > 0 favors the tested practice, < 0 favors the standard (comparison) practice).

(2) Cohen's *d* (*d*-score) is an estimate of the standardized mean difference between two practices when the underlying data are continuous. Many formulae exist to convert or transform reporting indices into Cohen's *d,* providing a common index on which to compare study results. The resulting effect size centers on zero and has the following interpretation:

- *d*-score > 0: new practice is more successful than standard practice
- *d*-score = 0: no differences between new practice and standard practice
- *d*-score < 0: new practice is less successful than standard practice

The further the *d*-score is from zero the more successful the practice is relative to the comparison practice when positive and the less successful when negative.

3.3.5 RATING INDIVIDUAL STUDY QUALITY

The evidence summary format is designed to provide the relevant content corresponding to the evaluation methods piloted in Phase 2 for rating individual study quality using four dimensions listed below. If all four dimensions receive the maximum number of points, the overall study quality rating for an individual study would be a "10". Principles for making judgments and guidance on each of the rating criteria, including specific reasons for deducting points from the maximum, are provided for each dimension in the Guide to Rating Study Quality in Appendix F.

Study (maximum of 3 points)

[6] See Appendix G for the detailed formulas used to calculate effect sizes.

- Study design
- Facility / setting
- Time period
- Sampling limitations (selection biases)
- Appropriateness of comparator

Practice (2 points maximum)
- Description
- Duration
- Requirements (equipment, staff, training, costs)

Outcome measure(s) (2 points maximum)
- Description, relevance, and validity
- Recording method reliability

Findings/Results (3 points maximum)
- Type of findings
- Findings/effect size
- Potential biases (uncontrolled deviations and results/conclusions bias)

For each individual study concerning a particular practice, these dimensions can be arrayed in a summary table like the following:

Practice A	Study Characteristics (3 points)	Practice Characteristics (2 points)	Outcome Measures (2 points)	Results/ Other (3 points)	Overall Study Quality Rating (2)
Study 1					
Study 2					
Study 3					
…					
Study N					

This 10-point scale supports the following categorical study quality ratings
- Good: 8-10 points total (all four dimensions)
- Fair: 5-7 points total
- Poor: ≤ 4 points total

A “poor” quality rating indicates a study has significant flaws, implying biases that may invalidate results. Thus, individual studies with a “poor” quality rating are excluded from consideration as evidence.

3.4 EVALUATION METHODS AND USE OF EXPERT PANELS

With the published and unpublished evidence collected, screened, abstracted, standardized and summarized by the CDC/Battelle LMBP Review Team, responsibility for completing the evaluation of the aggregate body of evidence was assigned to multidisciplinary Expert Panels selected for each review topic (see Appendix B for each panel’s roster). The LMBP Expert Panels were asked to review the standardized practice evidence summary tables, individual study ratings, and forest plot figures for each study documenting the effectiveness of practices associated with their panel’s topic area. They

used this information to reach consensus ratings for effect size, overall consistency and overall strength of evidence ratings. From their evidence evaluations, the Expert Panels were then asked to draft an evidence-based recommendation regarding the adoption of the practice. The practice-specific evidence reviews, evaluations and draft recommendations for each practice were then reviewed by the Workgroup in their capacity as the pilot test recommending body.

The Expert Panels included subject matter experts in the topic area, as well as experts in evidence review methods and in laboratory management. Experts were identified based on their publication record as well as involvement and leadership in relevant organizations and initiatives, particularly those considered key stakeholders for laboratory practice recommendations. In addition, for the purposes of the pilot test, like other evidence-based recommending organizations' methods, experts among the Workgroup were included as panelists to ensure support and continuity between the work of the Expert Panel and the Workgroup. By inviting individuals with expertise that were also associated with laboratory and professional organizations to serve as Expert Panelists, another objective was to increase and broaden the participant-observers in this stage of the pilot test. This facilitated making the development and testing of the methods transparent and accessible to a wider audience that can provide useful feedback about refinements that will benefit implementation planning for the evidence review process.

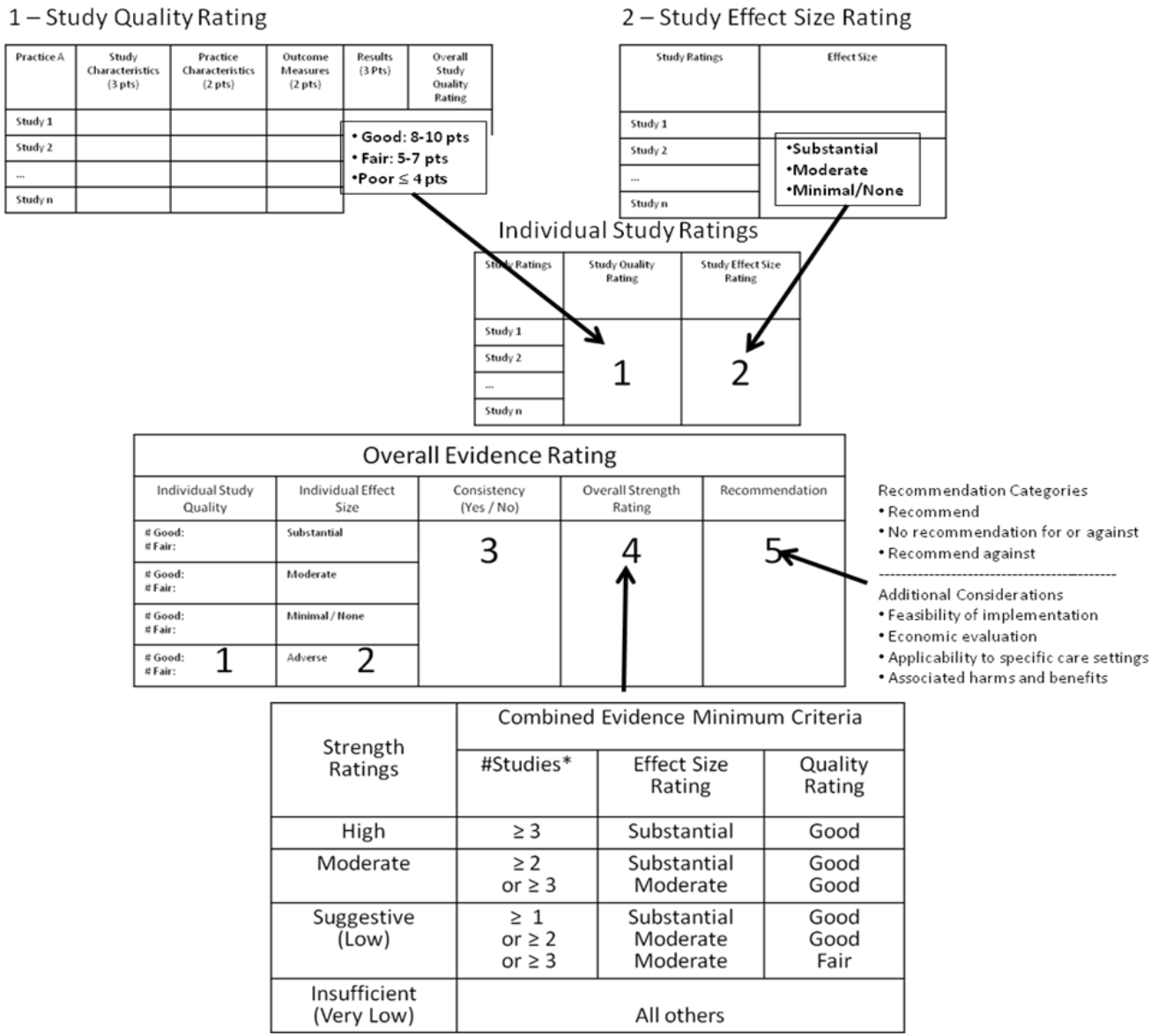
3.5 STEP 4: ANALYZE – RATE THE BODY OF EVIDENCE

The aggregate body of evidence generated in Step 3. APPRAISE was analyzed after the abstracted and standardized information for individual studies were entered into a practice's Body of Evidence Table which includes each studies quality ratings across four quality dimensions (study characteristics, practice characteristics, outcome measures used and results observed). Figure 6 provides a schematic of the overall approach that was used to analyze evidence from both the published and unpublished sources for a single practice. The approach involves four main steps leading to one of three practice implementation recommendations (i.e., for, against, and no recommendation for or against):

1. Rating **individual study quality** (good, fair, poor), based on evaluating four dimensions (using a 10-point scale)
    a. Study characteristics
    b. Practice characteristics
    c. Measure(s) used
    d. Result(s) observed

2. Rating the observed **individual study effect size(s)** categorically on magnitude (substantial, moderate, minimal/none) and relevance to the review question (direct, less direct, indirect)

FIGURE 6. INDIVIDUAL STUDY QUALITY AND EFFECT SIZE RATINGS ARE TRANSLATED INTO AN OVERALL RATING FOR EVIDENCE OF EFFECTIVENESS AND PROVIDE THE BASIS FOR A BEST PRACTICE RECOMMENDATION



Strength Ratings	Combined Evidence Minimum Criteria		
	#Studies*	Effect Size Rating	Quality Rating
High	≥ 3	Substantial	Good
Moderate	≥ 2 or ≥ 3	Substantial Moderate	Good Good
Suggestive (Low)	≥ 1 or ≥ 2 or ≥ 3	Substantial Moderate Moderate	Good Good Fair
Insufficient (Very Low)	All others		

3. Assessing the **consistency of all studies'** (body of evidence) observed effect sizes based on direction and magnitude.

4. Rating the overall **strength of a body of evidence** based on the total number of studies by their quality ratings and effect size ratings.

Detailed guidance was provided to the Expert Panelists on how to characterize individual study quality according to the four analytical dimensions listed above (see Appendix F).

3.5.1 EFFECT SIZE RATINGS

Expert Panel members were asked to confirm the summary judgment for each observed effect size for each individual study in one of three categories: "Substantial," "Moderate," "Minimal/None." In Phase 2, it was assumed that because these ratings are specific to topic areas, Expert Panel input would be necessary for specifying the value ranges associated with each category for the relevant outcome measures. In practice, this approach proved unwieldy as there are not necessarily evidence-based or otherwise available standards for estimating a clinically relevant impact of laboratory medicine pre- and post-analytic practices associated with a given topic area. Therefore, meta-analytic graphical displays (forest plots) of effect size magnitude and the 95% confidence intervals for that point estimate were adopted in Phase 3 and were used to make effect size rating decisions. In general, magnitude of the effect size was used to determine if the effect size was substantial, moderate, or minimal/none. The general guidelines for making this determination was: if the confidence interval did not include null (1 if logged odds ratios or *d* scores, 0 if odds ratio), then the finding was considered to be 'substantial;' If the confidence interval included zero, but the probability of impact was substantially positive, then the finding was considered to be 'moderate;' effect sizes that centered on or near zero were considered 'minimal/none.' An example of an effect size rating graph is provided in Figure 7.

3.5.2 CONSISTENCY RATING

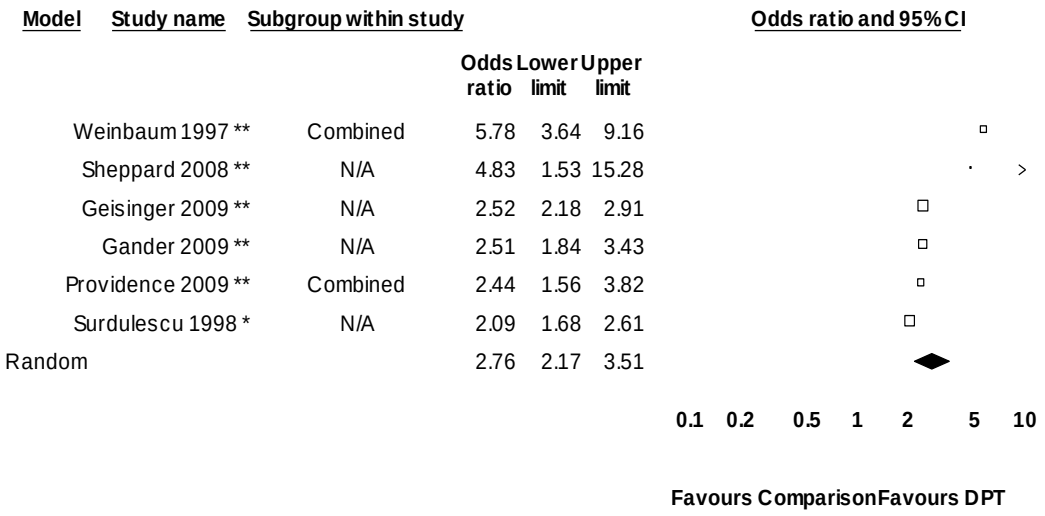
As established by AHRQ (2007), consistency across individual studies for a given practice is measured as a dichotomous variable (i.e., "consistent" or "not consistent") based on similarity in reported effect sizes from studies included in a body of evidence for a given practice. A body of evidence for a given practice is considered "consistent" if the evidence is all in the same direction and within a reasonably narrow range. For the evaluation methods, "reasonability" is determined by consensus expert judgment as informed by the effect size meta-analysis results and graphic representation (forest plot).

3.5.3 OVERALL STRENGTH OF A BODY OF EVIDENCE

Four overall strength rating categories were established: "High," "Moderate," "Suggestive (Low)," and "Insufficient". Initially, these rating categories were defined in terms derived from Guyatt et al. (2008), which expressed the strength ratings in terms of how likely it is that additional evidence would change the confidence in the direction and general magnitude of the observed effect. The LMBP Workgroup recommended that the

category definitions be changed to reflect the quality of the evidence and effect size observed, rather than attempting to anticipate the impact of future potential evidence.

FIGURE 7: EXAMPLE OF AN EFFECT SIZE RATING GRAPH: DEDICATED PHLEBOTOMY TEAMS



The revised definitions for these categories, modeled after the US Preventive Services Task Force (2008) are as follows:

High: An adequate volume of evidence is available and includes consistent evidence of substantial healthcare quality changes from studies without major limitations.

Moderate: Some evidence is available and includes consistent evidence of substantial healthcare quality changes from without major limitations.

Suggestive: Limited evidence is available and includes consistent evidence of moderate healthcare quality changes from a small number of studies without major limitations; OR the quality of some of the studies' design and/or conduct is limited.

Insufficient: Any estimate of an effect is very uncertain. Available evidence of effectiveness is:

- Inconsistent or weak; OR
- Consistent but with a minimal effect; OR
- Contained in an inadequate volume to determine effectiveness

As shown in Table 1 below, in Phase 3 the overall strength rating was derived from a minimum required number of studies with the various categorical ratings for individual study quality and effect size assigned by consensus of the Expert Panelists. Note that the "High," "Moderate," and "Suggestive" effect size ratings require "consistency" across studies included in a practice evidence base as discussed above. For example, to

achieve an overall strength rating of “high,” a body of evidence would have to include at least three studies with a quality rating of “good” that also have an effect size rating of “substantial” for the relevant outcome measures. To achieve a “moderate” strength rating, a body of evidence would have to include EITHER (1) at least two studies with a quality rating of “good” that also have an effect size rating of “substantial,” OR (2) at least three studies with a quality rating of “good” and an effect size rating of “moderate.”

TABLE 1. OVERALL EVIDENCE OF EFFECTIVENESS STRENGTH RATING

Strength Ratings	Combined Evidence Minimum Criteria		
	# Studies	Effect Size Rating	Quality Rating
High	≥ 3	Substantial	Good
Moderate	≥ 2 ≥ 3	Substantial Moderate	Good Good
Suggestive (Low)	≥ 1 ≥ 2 ≥ 3	Substantial Moderate Moderate	Good Good Fair
Insufficient (Very Low)	All others		

3.5.4 ROLE OF EXPERT OPINION

While it is acknowledged that judgment is necessary for the evaluation and interpretation of all evidence regardless of quality, rating systems that classify “expert opinion” as a category of evidence create confusion. Simple evidence grading systems facilitate incorporating the expert judgment of multiple perspectives, ideally representing all relevant stakeholders, by applying detailed and explicit criteria for rating or grading evidence quality, effect size, net impact, and overall strength to make the judgments transparent (Guyatt et al. 2008). For the purposes of the LMBP Phase 3 practice evidence reviews, “expert opinion” about the effectiveness of a given practice is not considered evidence, and is excluded from a practice’s evidence of effectiveness; only measured effectiveness is explicitly considered as evidence. The LMBP methods in their entirety and the exclusion from evidence of expert opinion of practice effectiveness in particular, are consistent with making the LMBP evidence review findings, and hence recommendations based on those findings, satisfying a transparent standard of being “evidence-based.” It is acknowledged that there may be some instances in which systematic measures are simply unavailable, in which case structured interpretive decision-making models (e.g., the Delphi process) may be used (Thomson et al. 2009). However, the results of applying such models are qualitatively different from evidence based on systematically organized observations and measurements.

3.6 METHODS FOR BEST PRACTICE RECOMMENDATIONS & ADDITIONAL CONSIDERATIONS

Each overall LMBP rating category for the overall strength of a body of evidence translates into one of three rating categories. In Phase 2, the LMBP Methods Development Team initially proposed four categories of recommendations: "*Strongly recommend*," "*Recommend*," "*No recommendation for or against*," and "*Recommend against*." However, it was found in practice that based on available evidence from both published and unpublished sources, it is difficult to distinguish between the first two ("*Strongly recommend*" and "*Recommend*"). It was also found that a number of different conditions could result in a "*No recommendation for or against*" finding (e.g., too few studies with substantial or moderate effect size, or the available studies are of relatively modest quantity and/or quality, or the available studies consistently show an effect size of 'minimal/none'). To make the recommendations more useful and informative, and less ambiguous, the LMBP Workgroup recommended further refinements to simplify the recommendation categories by limiting the number to three, with constructive explanations detailing multiple categorical yet distinct reasons supporting its application.

3.6.1 RECOMMENDATION CATEGORIES

Following the Workgroup's guidance, the recommendation rating categories, consistent with the GRADE group findings (2008), reflect the extent to which the available evidence gives one confidence that a practice will do more good than harm:

Recommend: The practice should be identified as a best practice for implementation in appropriate care settings, taking into account variations in implementation and/or care settings. This recommendation results from consistent and high or moderate overall evidence of effectiveness strength rating of desirable impacts.

No recommendation for or against: The potentially favorable impact on care outcomes and/or error reduction is not sufficient, or not sufficiently supported by evidence to indicate that it should be identified as a best practice for implementation in appropriate care settings. Additional studies may be warranted to strengthen the relevant evidence base. This recommendation results from insufficient evidence to determine effectiveness.

Recommend against: The practice should not be identified as a best practice for implementation because of consistent evidence of adverse effects.

3.6.2 ADDITIONAL CONSIDERATIONS

An important distinction can be made between evidence of effectiveness and evidence concerning other aspects of implementation, such as feasibility, costs, applicability to specific care settings, and other harms and benefits. In Phase 2 if studies were eligible for inclusion in the systematic review, implementation-related evidence was included in the Recommendation when it was available. In Phase 3, however, these additional considerations were not included in the Recommendation as it was observed that the evidence review did not systematically search for or record evidence on these additional

considerations beyond what was documented in the effectiveness literature, It is recommended in future reviews that such evidence be systematically included and incorporated in practice recommendations. Such evidence may be reported as:

- ***Feasibility of Implementation**:* Whether the practice is in current use and available for immediate application, whether it is able to be used in a variety of inpatient and/or outpatient settings, and whether significant barriers to implementation have been identified.

- ***Economic Evaluation**:* The cost of implementing a practice, the savings that are achieved with implementation, and the results of any cost-effectiveness and/or cost-benefit assessments that have been completed.

- ***Applicability to Specific Care Settings***: Whether the practice is suitable for use across a range of inpatient and outpatient care settings, targeted for point-of-care testing, or other information relevant to the practice's applicability.

- ***Associated Harms and Benefits***: Whether implementing the practice has had observed impacts for patient satisfaction, provider satisfaction, ability to measure and monitor quality and process improvement, standardization of protocols across a healthcare network or system, or other outcomes that contribute to improvements in patient safety and healthcare quality.

3.7 STEP 5: APPLY THE FINDINGS

Step 5 of the LMBP methods refers to the dissemination of systematic review findings for translation of evidence-based findings into practice. Dissemination plans for LMBP findings include peer-reviewed publications, press releases and presentations at scientific and professional meetings. Access to LMBP evidence review summaries will also be available online.

4.0 EVIDENCE REVIEW RESULTS

Preliminary evidence reviews were completed in Phase 3 for Patient Specimen Identification, Critical Values Reporting and Communication, and Blood Culture Contamination. The following discussion summarizes the recommendations and preliminary evidence review findings for each practice meeting the review inclusion criteria, while Appendix E provides a more detailed evidence summary table for all studies included in the preliminary evidence review.

4.1 PATIENT SPECIMEN IDENTIFICATION

As noted earlier, practices associated with this topic area are designed to reduce patient specimen identification errors and assure positive patient specimen identification. Practices for which enough evidence was available from unpublished and published sources to be included in the evidence review are:

- **Point-of-Care Testing Bar Coding Systems** - Automated patient and sample/test result identification system using bar-coded patient armbands and bar code scanners when using a testing device at or close to the patient.

- **Bar Coding Systems** - Electronic bar-coding on both patient and specimen used to establish positive identification of specimen as belonging to patient. This involves the use of scanners and capability to print labels.

Table 2 summarizes the preliminary findings from the evidence review, including best practice recommendations.

TABLE 2. EVIDENCE-BASED PRACTICE RECOMMENDATIONS: PATIENT SPECIMEN/TEST RESULT IDENTIFICATION

Practice	Recommendation Statement	Preliminary Findings
Point-of-Care Testing Bar Coding Systems *(3 published studies and 2 unpublished studies)*	The Laboratory Medicine Best Practices Workgroup recommends identifying the use of a bar coding process to consistently link patients and their specimen test results in point-of-care testing settings as a best practice to reduce or eliminate Patient Specimen/Test Result Identification errors. This is based on the strength of evidence for this practice and consistency of observed effects	An adequate volume of evidence is available and includes consistent evidence of substantial healthcare and safety changes from studies without major limitations. (5 studies, Odds Ratio =6.55, 95% CI 3.1-14.0).
Bar Coding Systems *(4 published studies and 4 unpublished studies)*	The Laboratory Medicine Best Practices Workgroup recommends identifying the use of a bar coding process to consistently link patients and their specimen through the entire testing process as a best practice to reduce or eliminate Patient Specimen Identification errors. This is based on the strength of evidence for this practice and consistency of observed effects	An adequate volume of evidence is available and includes consistent evidence of substantial healthcare and safety changes from studies without major limitations. (8 studies, log odds ratio = 2.45, 95% CI 1.6-3.3).

4.2 CRITICAL VALUES REPORTING AND COMMUNICATION

As noted earlier, practices associated with this topic area are designed to assure timely and accurate communication of critical value laboratory test results to the licensed responsible caregiver who can act on these results. Practices for which enough evidence was available from unpublished and published sources to be included in the evidence review are:

- **Automated Notification** – Automated alerting system or computerized reminders using mobile phones, pagers, email or other personal electronic devices to alert clinician of critical laboratory test results.

- **Call Center** – Critical value notification process centralized in a unit responsible for communication of critical value laboratory test results to the licensed caregiver.

Table 3 summarizes the preliminary findings from the evidence review, including best practice recommendations.

TABLE 3. EVIDENCE-BASED PRACTICE RECOMMENDATIONS: CRITICAL VALUE TEST RESULT COMMUNICATION

Practice	Recommendation Statement	Preliminary Findings
Automated Notification *(3 published studies)*	**No Recommendation for or Against**: The Laboratory Medicine Best Practices Workgroup cannot recommend for or against identifying the use of automated notification as a best practice to improve the timeliness and accuracy of critical value reporting due to limited number of studies with good quality rating that would be necessary to make a recommendation.	An insufficient volume of evidence is available of adequate quality showing a moderate or substantial improvement attributable to using this practice. (3 studies, Standard difference of means = 0.51, 95% CI -0.4 – 1.4)
Call Center *(1 published study, 2 unpublished studies)*	**No Recommendation for or Against**: The Laboratory Medicine Best Practices Workgroup cannot recommend for or against identifying the use of Call Centers as a best practice to improve the timeliness and accuracy of critical value reporting due to limited number of studies with good quality rating that would be necessary to make a recommendation.	An insufficient volume of evidence is available of adequate quality showing a moderate or substantial improvement attributable to using this practice. (3 studies, Standard difference of means = 0.81, 95% CI -0.52 – 2.15)

4.3 BLOOD CULTURE CONTAMINATION

As noted earlier, practices associated with this topic area are designed to reduce blood culture contamination. Practices for which enough evidence was available from unpublished and published sources to be included in the evidence review are:

- Dedicated Phlebotomy Teams: Use of certified phlebotomists to draw blood specimens for laboratory tests.
- Pre-packaged Prep Kits: Pre-packaged aseptic supplies that are prepared in-house or commercially purchased.
- Venipuncture vs. Intravenous catheter collection: Puncture of a vein through the skin *vs.* use of a thin flexible tube inserted into the body to withdraw blood for analysis.

Table 4 summarizes the preliminary findings from the evidence review, including best practice recommendations.

TABLE 4. EVIDENCE-BASED PRACTICE RECOMMENDATIONS: BLOOD CULTURE CONTAMINATION

Practice	Recommendation Statement	Preliminary Findings
Venipuncture *(7 published studies)*	**Recommend**: The Laboratory Medicine Best Practices Workgroup recommends identifying the use of venipuncture as the preferred technique for sample collection where the option exists in the clinical setting as a best practice to reduce or eliminate blood culture contamination. This is based on the strength of evidence for this practice and consistency of observed effects	An adequate volume of evidence is available and includes consistent evidence of substantial healthcare and safety changes from studies without major limitations. (7 studies, Odds Ratio = 2.63, 95% CI 1.85-3.72)
Dedicated Phlebotomy Teams *(4 published studies, 2 unpublished studies)*	**Recommend**: The Laboratory Medicine Best Practices Workgroup recommends identifying the use of dedicated phlebotomy teams for sample collection where the option exist in the clinical setting as a best practice to reduce or eliminate blood culture contamination. This is based on the strength of evidence for this practice and consistency of observed effects	An adequate volume of evidence is available and includes consistent evidence of substantial healthcare and safety changes from studies: 6 studies without major limitations. (Odds Ratio = 2.76, 95% CI 2.2-3.5)
Pre-packaged Prep Kits (*4 published studies*)	**No recommendation for or against**: The Laboratory Medicine Best Practices Workgroup cannot recommend for or against identifying the use of Commercial Prep Kits as a best practice to reduce blood culture contamination due to the limited improvement in effectiveness observed among the reported evidence.	An adequate volume of evidence is available and includes consistent evidence of minimal healthcare and safety changes from studies without major limitations. (4 studies, Odds Ratio =1.1, 95% CI 0.99-1.41)

4.3 DISCUSSION

The preliminary evidence reviews resulted in recommendations for identifying as “best practices” two practices for Patient Specimen Identification, and two practices associated with reducing Blood Culture Contamination. No recommendation for or against one practice associated with reducing blood culture contamination based on an adequate amount of evidence suggesting this practice is not observed to represent a substantial improvement over standard practices. No recommendation for or against identifying two practices associated with critical values reporting could be reached due to insufficient evidence available. Plans for moving the systematic review process towards ongoing implementation will tackle directly an enhanced approach to collecting relevant evidence from unpublished sources, and will carry forward the practices from Phase 3 to

complete the evidence reviews for publication. Only limited work to date has been completed on developing a process for reviewing implementation-related evidence ("additional considerations" in the Recommendation write-up). Plans for moving the review process forward will also need to attend to enhancing and /or limiting the review of available evidence or information related to practical aspects as well.

5.0 CONCLUSIONS

Evidence review and evaluation methods developed and tested during this phase were substantially refined from the simplified approach that had been formulated in earlier project phases. Efforts to recruit facilities and healthcare organizations to participate in a network to provide unpublished evidence provided helpful insight into the factors that constrain and encourage participation. Phase 3 network efforts have demonstrated that recruitment success depends upon several critical factors, including contacts with facility representatives who are both knowledgeable and invested with appropriate decision-making authority; appropriate formal letters of invitation and endorsement; information that meets the needs of relevant IRB chairs and other administrative review offices; assurances of confidentiality when requested. Characterization of the roles and responsibilities of the Workgroup, Expert Review Panels, and the staff support team evolved over the course of the Pilot Phase, helping CDC to further specify organizational requirements that must be met to support systematic evidence reviews and the production of best practice recommendations on an ongoing basis. Validated LMBP systematic review methods are expected to be published in the peer-review literature.

In moving towards sustained implementation, it is recommended that refinements and enhancements in the systematic evidence review process include some or all of the following activities.

5.1 METHODS: TOPIC AREA SELECTION

Refine and standardize the process by which systematic review topic areas are selected and associated candidate practices are nominated. Topic selection criteria established early in the Initiative's development still apply (burden of problem/quality gap; preventability, availability of existing knowledge, potential effectiveness, operational management, and potential economic benefit), but further refinements are needed in soliciting and responding to suggestions from the field.

5.2 METHODS: ANALYTIC FRAMEWORK

Refine and standardize methods for schematic representation of a topic area analytic framework for each review question including:

- Formalize a process for establishing functional requirements for practices associated with a selected topic area. A "process mapping" approach may help to outline work flows and common points of intervention at which practices can achieve improvements in healthcare quality outcomes.
- Identify processes from domains of application outside of laboratory medicine that meet the same functional requirements, increasing the likelihood that

evidence of effectiveness from these other domains will be regarded as relevant to laboratory medicine practices.

5.3 METHODS: SEARCH, SCREENING AND DATA ABSTRACTION METHODS

Make further improvements to the review methods and electronic data abstraction tool including:

- Refine, standardize, and document literature search strategy to generate relevant published materials in a broader array of journals and published conference proceedings.
- Develop standardized search and reporting functions for reference and study databases
- Improve guidance and standardization for screening and abstraction methods for reviewers
- Refine reviewer/user interface enhancements for data abstraction
- Structure and formatting of data abstraction template more directly linked with evidence summary templates and individual study evaluation criteria,
- Further standardization of outcome measures, definitions, and their categorization to minimize topic area-specific programming and maximize comparability
- Develop and implement standardized methods for screening and capturing non-effectiveness evidence related to feasibility of implementation, applicability, economic evaluation and harms and benefits and/or other newly developed criteria.

5.4 METHODS: EVIDENCE SUMMARY AND EVALUATION

- Finalize evidence summary presentation formats along with development of standardized content and terms to facilitate and ensure consistent evaluations, and when applicable statistical meta-analyses, and recommendation statements (for the LMBP topic area Expert Panels and Workgroup), and for publishing and disseminating evidence reviews and evidence-based recommendations.

- Specify methods for including, evaluating and synthesizing additional non-effectiveness evidence related to implementation feasibility, economic evaluation, applicability (settings, populations, contextual variables) and harms and benefits, incorporating concepts of external validity and internal validity.

- Further refine protocols for nominating, selecting, and guiding the work of expert panelists so that panelists have a clear idea of their roles and responsibilities relative to the Recommending Body and support staff, and panel composition is adequately diversified to represent key stakeholders' perspectives to produce unbiased and scientific evidence reviews.

- Further refine protocols for guiding the work of the LMBP Workgroup (or if not overlapping a Recommending Body) so that members of this body have a clear idea of their roles and responsibilities relative to the expert panelists and support staff.

5.5 NETWORK DEVELOPMENT FOR UNPUBLISHED EVIDENCE

- Further develop the network as the principal source for unpublished evidence. Expanding and maintaining this network is essential to the future sustainability of an evidence-based laboratory medicine practice recommendations process, as the main challenge to its success remains insufficient published evidence.

- Further refine guidance to network participants on informational requirements for submitting evidence.

- Develop and implement an education / curriculum strategy that familiarizes laboratory managers with methods for improving the quality of unpublished process improvement / quality assurance studies so that data from these studies are consistently available to inform “best practice” recommendations.

- Expand strategies to extend the breadth and depth of the network to provide greater opportunities for identifying participating organizations and individuals within those organizations responsible for relevant practice evaluations and quality improvement initiatives.

- Maintain a network tracking database with strategic information to facilitate contacts, targeted follow-up as well as routine communication with network affiliates.

5.6 ORGANIZATIONAL DEVELOPMENT AND SUSTAINABILITY

- Create a specific business plan for implementation and funding alternative models based on collaboration with key stakeholders

- Develop and implement communication, publication and other dissemination strategies based on collaboration with key stakeholders to optimize impact of evidence reviews and further the implementation of evidence-based methods and standards for quality improvement in laboratory medicine.

- Develop a process for assuring a pipeline of future topic areas and priorities for evidence reviews based on broad stakeholder engagement, including identification of appropriate evidence.

6.0 REFERENCES CITED

AHRQ, *Methods Reference Guide for Effectiveness and Comparative Effectiveness Reviews* (Version 1.0), October 2007. http://effectivehealthcare.ahrq.gov/repFiles/2007_10DraftMethodsGuide.pdf (accessed January 13, 2009)

Askeland RW, McGrane S, Levitt JS, Dane SK, Greene DL, Vandeberg JA, Walker K, Porcella A, Herwaldt LA, Carmen LT, and Kemp JD. 2008. Improving transfusion safety: implementation of a comprehensive computerized bar code-based tracking system for detecting and preventing errors. *Transfusion*. 48(7):1308-17.

Barenfanger J, Sautter RL, Lang DL, Collins SM, Hacek DM, and Peterson LR. 2004. Improving patient safety by repeating (read-back) telephone reports of critical information. *Am J Clin Pathol.* 121(6):801-3.

Barr JT and Silver S. 1994. The total testing process and its implications for laboratory administration and education. *Clin Lab Manage Rev.* 8(5):526-42.

Bologna LJ, Lind C., and Riggs RC. 2002. Reducing major identification errors within a deployed phlebotomy process. *Clin Leadersh Manag Rev.* 16(1):22-6

Briss PA, Zaza S Pappaioanou M, Fielding J, Wright-DeAguero L, Truman BI, Hopkins DP, Mullen PD, Thompson RS, Woolf SH, Carande-Kulis VG, Anderson L, Hinman AR, McQueen DV, Teutsch SM, Harris, JR, and the Task Force on Community Preventive Services. 2000. Developing an Evidence-Based *Guide to Community Preventive Services*—Methods. *Am J Prev Med* 18(1S):25-43.

Chan JC, Chu RW, Young BW, Chan F, Chow CC, Pang WC, Chan C, Yeung SH, Chow PK, Lau J, and Leung PM. 2004. Use of an electronic barcode system for patient identification during blood transfusion: 3-year experience in a regional hospital. *Hong Kong Med J.* 10(3):166-71.

Clark P, Rennie I, and Rawlinson S. 2001. Quality improvement report: Effect of a formal education programme on safety of transfusions. *BMJ.* 323(7321):1118-20.

Colard, DR. 2004. Reduction of Patient ID Errors Using Technology. *Point of Care: The Journal of Near-Patient Testing & Technology*. 3(2):84.

Committee on the National Quality Report on Health Care Delivery Margarita P. Hurtado, Elaine K. Swift, and Janet M. Corrigan, Editors. *Envisioning the National Health Care Quality Report*. Board on Health Care Services, Institute Of Medicine, National Academy Press, Washington, D.C., March 2001.

Coyle GA and Heinen M. 2002. Scan your way to a comprehensive electronic medical record. Augment medication administration accuracy and increase documentation efficiency with bar coding technology. *Nurs Manage*. 33(12):56, 58-9.

Davies A, Staves J, Kay J, Casbard A, and Murphy MF. 2006. End-to-end electronic control of the hospital transfusion process to increase the safety of blood transfusion: strengths and weaknesses. *Transfusion* 46(3):352-64.

Diconsiglio J. 2006. System ensures fluid specimen collection. *Mater Manag Health Care*. 15(1):36-8.

Dock B. 2005. Improving the accuracy of specimen labeling. *Clin Lab Sci*. 18(4):210-2.

Grzybicki DM, Turcsanyi B, Becich MJ, Gupta D, Gilbertson JR., and Raab SS. 2005. Database construction for improving patient safety by examining pathology errors. *Am J Clin Pathol*. 124(4):500-9.

Guyatt GH et al. 2008. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336: 924-26.

Hayden RT, Patterson DJ, Jay DW, Cross C, Dotson P, Possel RE, Srivastava DK, Mirro J., and Shenep JL. 2008. Computer-assisted bar-coding system significantly reduces

clinical laboratory specimen identification errors in a pediatric oncology hospital. *J Pediatr.* 152(2):219-24.

Horvath AR and Pewsner D. 2004. Systematic reviews in laboratory medicine: principles, processes and practical considerations. *Clinica Chimica Acta* 342:23–39.

Institute of Medicine. 2000. *To err is human: building a safer health system*. Washington, DC: National Academy Press.

Institute of Medicine. 2001. *Crossing the quality chasm: a new health care system for the 21st century*. Washington, DC: National Academy Press.

Jenkins JJ, Mac Crawford J, and Bissell MG. 2007. Studying critical values: adverse event identification following a critical laboratory values study at the Ohio State University Medical Center. *Am J Clin Pathol.* 128(4):604-9.

Johnson VR, Hummel J, Kinninger T, and Lewis RR. 2004. Immediate steps toward patient safety. *Healthc Financ Manage*. 58(2):56-61.

Khan KS, ter Riet G, Glanville J, Sowden AJ, Kleijnen J, (eds). 2001. *Undertaking Systematic Reviews of Research on Effectiveness: CRD's Guidance for those Carrying Out or Commissioning Reviews*. CRD Report No. 4 (2nd edition). York (UK): National Health Service Centre for Reviews and Dissemination. March, 2001.

Killeen JP, Chan TC, Jones K, and Guss DA. 2005. Impact of Bar Coding Technology and Computerized Physician Order Entry on Reducing Laboratory Specimen Misidentification Errors in the Emergency Department. *Acad Emerg Med* 12(5), Suppl. 1: 123.

Kuperman GJ, Boyle D, Jha A, Rittenberg E, Ma'Luf N, Tanasijevic MJ, Teich JM, Winkelman J., and Bates DW. 1998. How promptly are inpatients treated for critical laboratory results? *J Am Med Inform Assoc*. 5(1):112-9.

Kuperman GJ, Teich JM, Tanasijevic MJ, Ma'Luf N, Rittenberg E, Jha A, Fiskio J, Winkelman J, and Bates DW. 1999. Improving response to critical laboratory results with automation: results of a randomized controlled trial. *J Am Med Inform Assoc.* 6(6):512-22.

Nichols JH, Bartholomew C, Brunton M, Cintron C, Elliott S, McGirr J, Morsi D, Scott S, Seipel J, and Sinha D. 2004. Reducing medical errors through barcoding at the point of care. *Clin Leadersh Manag Rev*. 18(6):328-34.

Park HI, Min WK, Lee W, Park H, Park CJ, Chi HS, and Chun S. 2008. Evaluating the short message service alerting system for critical value notification via PDA telephones. *Ann Clin Lab Sci*. 38(2):149-56.

Price CP, Glenn JL, and Christenson RH, 2009. *Applying Evidence-Based Laboratory Medicine: A Step-By-Step Guide.* Washington, DC: American Association for Clinical Chemistry.

Quillen K, and Murphy K. 2006. Quality improvement to decrease specimen mislabeling in transfusion medicine. *Arch Pathol Lab Med*. 130(8):1196-8.

Rao AC; Burke DA, and Dighe AS. 2005. Implementation of Bar Coded Wristbands in a Large Academic Medical Center: Impact on Point of Care Error Rates. *Point of Care: The Journal of Near-Patient Testing & Technology*. 4(3):119-122.

Sandler SG, Langeberg A, and Dohnalek L. 2005. Bar code technology improves positive patient identification and transfusion safety. *Dev Biol* 120:19-24.

Saxena S, Kempf R, Wilcox S, Shulman IA, Wong L, Cunningham G, Vega E., and Hall S. 2005. Critical laboratory value notification: a failure mode effects and criticality analysis. *Jt Comm J Qual Patient Saf*. 31(9):495-506.

Shabot MM, LoBue M, and Chen J. 2000. Wireless clinical alerts for physiologic, laboratory and medication data. *Proc AMIA Symp*. 789-93.

Shaneyfelt T, Baum KD, Bell D et al. Instruments for Evaluating Education in Evidence-Based Practice: A Systematic Review. *JAMA*. 2006;296(9):1116-1127.

Simpson JB. 2001. A unique approach for reducing specimen labeling errors: combining marketing techniques with performance improvement. *Clin Leadersh Manag Rev*. 15(6):401-5.

Tate KE, Gardner RM, and Scherting K. 1995. Nurses, pagers, and patient-specific criteria: three keys to improved critical value reporting. *Proc Annu Symp Comput Appl Med Care.* 164-8.

Thomson R; Lewalle P; Sherman H, Hibbert P; Runciman W, and Castro G. 2009. Towards an International Classification for Patient Safety: a Delphi survey. *International Journal for Quality in Health Care* 21(1):9-17

Turner CL, Casbard AC, and Murphy MF. 2003. Barcode technology: its role in increasing the safety of blood transfusion. *Transfusion*. 43(9):1200-9.

US Preventive Services Task Force, Grade Definitions. November 2008. Agency for Healthcare Research and Quality. http://www.ahrq.gov/clinic/uspstf/grades.htm (accessed January 13, 2009).

Valenstein PN, Raab SS, and Walsh MK. 2006. Identification errors involving clinical laboratories: a College of American Pathologists Q-Probes study of patient and specimen identification errors at 120 institutions. *Arch Pathol Lab Med.* 130(8):1106-13.

Zaza S; Briss PA, and Harris KW, (eds). 2005. *The guide to community preventive services: What works to promote health? Task Force on Community Preventive Services.* New York: Oxford University Press.

APPENDIX A. LABORATORY MEDICINE BEST PRACTICES WORKGROUP ROSTER (2009)

Raj Behal, MD, MPH
Associate Chief Medical Officer
Senior Patient Safety Officer
Rush University Medical Center

Robert H. Christenson, PhD, DABCC, FACB
Professor of Pathology and Medical and Research Technology, University of Maryland Medical Center

John Fontanesi, PhD
Director, Center for Management Science in Health; Professor of Pediatrics and Family and Preventive Medicine
University of California, San Diego

Julie Gayken, MT(ASCP)
Director of Laboratory Services
Anatomic & Clinical Pathology
Regions Hospital

Cyril (Kim) Hetsko, MD, FACP
Clinical Professor of Medicine, University of Wisconsin-Madison
Chief Medical Officer, COLA
Trustee, American Medical Association

Lee Hilborne, MD, MPH
Professor of Pathology and Laboratory Medicine, UCLA David Geffen School of Medicine, Center for Patient Safety and Quality

James Nichols, PhD
Director, Clinical Chemistry
Department of Pathology
Baystate Medical Center

Mary Nix, MS, MT(ASCP)SBB
Project Officer, National Guideline Clearinghouse; National Quality Measures Clearinghouse; Quality Tools; Innovations Clearinghouse
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephen Raab, MD
Director, Cytopathology Laboratory
University of Colorado Cancer Center

Ann Vannier, MD
Director, Southern California Regional Reference Laboratory
Kaiser-Permanente Healthcare Systems

Ann Watt, MBA, RHIA
Project Director
Group on Standardized Performance Measurement Division of Research
The Joint Commission

Sousan S. Altaie, PhD (*ex officio*)
Scientific Policy Advisor, Office of In Vitro Diagnostic Device (OIVD) Evaluation and Safety Center for Devices and Radiological Health (CDRH), FDA

James A. Cometa (*ex officio*)
Team Leader/Program Analyst
Centers for Medicare and Medicaid Services (CMS)

APPENDIX B. EVIDENCE PANEL ROSTERS

Expert Panel 1 – Practices Associated with Reducing Patient / Specimen ID Errors

- Stephen Raab, MD, PhD (U Colorado Cancer Center)
- James Nichols,PhD (Bay State Health Systems)
- Steve Kahn, PhD (Loyola Medical Center of Chicago)
- Paul Valenstein, MD, FCAP (St. Joseph Mercy Hospital)
- Denise Geiger, PhD (John T. Mather Hospital)
- David Hopkins, MD, MPH (CDC)
- Julie Gayken, MT(ASCP) (Regions Hospital)
- Ronald Schifman, MD, MPH (Tucson Veterans Administration Health Center)

Expert Panel 2 – Practices Associated with Communicating Critical Values

- Robert Christenson, PhD, DABCC, FACB (U Maryland Medical Center
- Corinne Fantz, PhD (Emory U)
- Dana Grzybicki, MD, PhD (U Colorado Cancer Center)
- Lee Hilborne, MD, MPH (RAND / UCLA Medical School)
- Kent Lewandrowski, MD, PhD (Harvard Medical School)
- Mary Nix, MS, MT(ASCP)SBB (AHRQ)
- Rick Panning, CLS(NCA), MBA (American Red Cross)

Expert Panel 3 – Practices Associated with Blood Culture Contamination

- Dennis Ernst, PhD (Center for Phlebotomy Education)
- Dana Grzybicki, MD, PhD (U Colorado Denver)
- Margret Oethinger, MD, PhD (Providence Portland Medical Center)

- Stephen Raab, Director, MD, PhD (U Colorado Cancer Center)
- Ronald Schifman, MD, MPH (Southern Arizona VA Healthcare System)
- Ann Vannier, MD (Kaiser-Permanente Healthcare Systems)
- Melvin Weinstein, MD(U of Medicine, Dentistry of New Jersey-Robert Wood Johnson Medical School)

APPENDIX C. ROLES AND RESPONSIBILITIES OF WORKGROUP AND EXPERT PANELISTS

The Laboratory Medicine Best Practices Workgroup

Composition: 13 invited members, and two ex officio representatives from federal agencies (CMS and FDA); members are clinicians, pathologists, laboratorians, and specialists in systematic evidence reviews

Main Functions: Provides overall guidance and feedback on developing review and evaluation methods for making evidence-based best practice recommendations. As the "Recommending Body" for the LMBP pilot test, the Workgroup reviews, provides guidance and makes recommendations on:

- evaluation methods for producing evidence-based best practices recommendations
- systematic evidence review methods used by Expert Panel and CDC/Battelle Review team to synthesize published and unpublished studies
- a format for evidence summaries and draft best practice recommendations prepared by Expert Panels and CDC /Battelle Review Team staff
- strategies and methods for presenting and disseminating recommendations
- strategies and alternatives for implementation of an organizational support for Laboratory Medicine Best Practices methods
- recruitment of Laboratory Medicine Best Practices Network affiliates for unpublished studies
- topic area selection and criteria for practice reviews

Expert Panels

Composition: Each topic area Expert Panel will have 7-9 panelists, including:

- 2-3 Work Group members with relevant topic area content expertise
- 2-3 topic area content experts who are <u>not</u> Work Group members
- 1 specialist in evidence review methods
- 2 specialists in laboratory management, including administrative and laboratorian specialties

Main Functions: The Expert Panels are assembled for the specific purpose of providing completing evidence reviews, although during the pilot test they are also asked to provide guidance and feedback on the systematic review methods for the collection and synthesis of published and unpublished evidence for a given topic area. The members of each panel will:

- review and provide guidance on pilot test strategy prepared by the CDC/Battelle Review team including topic area analytic framework and review questions, inclusion and exclusion criteria, search, abstraction, review and evaluation methods applied to the published and unpublished evidence
- select practices and formulate descriptions and parameters for the evidence review including identification of relevant outcome measures for evaluating practice effectiveness and feasibility of implementation with support from the CDC/Battelle Review Team staff
- review the preliminary practice evidence summaries developed using the evidence collection, screening, abstraction and review methods
- apply and provide feedback on evaluation methods developed by the CDC/Battelle Review Team to produce ratings for individual study quality and effect size
- reach consensus on topic area evidence review quality and effect size rating categories, consistent with the evidence summary format
- evaluate individual practices' overall strength of evidence, effect size consistency (i.e., direction and magnitude)
- develop final draft practice evidence summaries and draft recommendations to be presented to the LMBP Workgroup
- provide feedback to the Review Team concerning limitations of the evidence review process and possible improvements

CDC / Battelle Team

***Composition*:** CDC Project Director, Task Leader for LMBP Methods Development, Task Leader for Network Development and Organizational Requirements, evidence review specialists, communications specialist – each CDC role has a Battelle counterpart.

Main Functions: The Team's work spans several functions, including:

Methods Development

- developing protocols for identification and selection of topics and associated practices
- developing analytic frameworks linking outcomes with practices and the healthcare quality issues / patient safety problems they are intended to address
- developing protocols for collecting published and unpublished practice-related evidence

- developing protocols for systematically collecting, reviewing and synthesizing published and unpublished evidence

- developing criteria for evaluating and rating practice-specific evidence including individual study quality, effect size, consistency of individual study findings, overall strength of a body of evidence, and other implementation considerations such as cost and feasibility

Literature Search

- Developing and presenting search strategies, review criteria and work plans to conduct a systematic literature search and review of search results

- Conducting the literature search and results review

- Summarizing and maintaining a record of the literature search results

Network Recruitment & Unpublished Evidence Harvest

- Identifying and contacting potential facilities and organizations willing to share relevant unpublished practice studies/assessments

- Providing sufficient information about the purpose and conduct of the evidence reviews to facilitate a network candidate's internal administrative review of CDC's invitation to participate

- Maintaining a complete database of recruitment efforts, including a record of recruitment outcomes

- Obtaining and cataloguing unpublished materials received

Evidence Screening and Review

- Applying screening (inclusion and exclusion) criteria to published search results and unpublished information received

- For unpublished evidence, obtaining additional information and clarification as needed when documentation is incomplete

- Maintaining an accurate record of screening dispositions

- Completing a standardized format summarizing information from published and unpublished sources

- Abstracting relevant information from document sources for potential inclusion in evidence review

Expert Panel Facilitation

The CDC/Battelle Team facilitates the work of each topic area Expert Panel:

- ensuring that the panelists have the materials they need to complete the evidence review
- facilitating Panel discussions
- preparing written summaries of Panel discussions
- drafting evidence summary, recommendations, and other supporting documentation
- providing other coordination activities as needed to promote a transparent, systematic, and objective review of available evidence

Workgroup Facilitation

The CDC/Battelle Team facilitates the Workgroup's involvement by:

- ensuring that the Workgroup members have the materials they need to provide timely feedback and recommendations on methods development and implementation strategy
- keeping the Workgroup apprised of the LMBP work plan and progress on methods development and pilot test implementation
- scheduling, coordinating, and preparing written summaries of regular Workgroup meetings

Organizational Development Planning

The CDC/Battelle Team is evaluating organizational development needs for implementing the LMBP evidence review and evaluation methods for making evidence-based recommendations

- Attributes and Features of Similar Efforts
- Organizational Structure and Governance
- Requirements for getting started
- Requirements for long-term sustainability
- Effective approaches to disseminating recommendations

Documentation and Dissemination

The CDC/Battelle Team is responsible for

- documenting the implementation and outcomes of the pilot test

- creating and disseminating information about the systematic evidence review and evaluation methods used and the evidence-based recommendations formulated
- receiving and summarizing feedback from key stakeholders about LMBP methods and best practice recommendations.

APPENDIX D. LITERATURE SEARCH STRATEGIES

Topic Area: Patient/Specimen Identification Errors

Purpose: To identify candidate practices for accurate patient/specimen identification and characterize specific process attributes.

Review Question: What are effective interventions/practices for reducing patient/specimen identification errors?

Search Strategy: Focus on identifying relevant published literature that categorizes/defines identification errors and/or identifies potential interventions/practices to reduce identification errors. Information sources are to be screened by the CDC Review Team to select articles/other literature for full-text review.

A comprehensive search of English language literature published during or after 1995 using multiple databases and other strategies. These include:

a. PubMed, MedLine, CINAHL, and Cochrane databases

b. Professional guidelines electronic databases (CLSI, ISO, NACB)

c. Hand searching journals of relevance to the review topic, reports, conference proceedings, and technical reports.

d. Reference lists of relevant reviews

e. Key informants: consultation with sub workgroup for relevant information sources

The title/abstract of the article should address:

a. A bar-coding technology implemented to establish positive identification of patients, and/or laboratory test orders and results,

b. Technology system components that interface with electronic medical and/or laboratory information systems,

c. Assessment or monitoring of bar-coding system,

d. Identified outcome measure.

At least two reviewers will independently screen the search results by applying exclusion criteria.

A pre-abstraction reference list of literature will be generated.

Initial Search Terms (results)

- Laboratory identification errors (187)*
- Identification errors AND patient AND specimen (25)***
- Laboratories AND identification systems AND specimen misidentification (0)*

- Specimen labeling errors (11)***
- Information systems AND hospitals AND reduce identification errors (4)***

*date/language exclusion criteria applied

**duplicate records removed

***both date/language exclusion and duplicate records removed

Additional Search Terms

- Patient specimen identification ((260)***
- Patient specimen identification errors (0)***
- Laboratory interventions to reduce identification errors (0)
- Laboratory methods to reduce identification errors (0)**
- Laboratory practices to reduce identification errors (0)**
- Laboratory strategies to reduce identification errors (0)**
- Reduce patient specimen identification errors (0)**
- Reducing patient specimen identification errors (0)**
- Practices for reducing patient specimen identification errors (0)**
- Strategies to reduce patient specimen identification errors (0)**
- Methods to reduce patient specimen identification errors (0)**
- Practices AND strategies to reduce patient specimen identification errors (0)
- Strategies to reduce identification errors (10)***
- Practices AND strategies to reduce identification errors (0)
- Practices to reduce identification errors (6)**
- Practices to reduce patient specimen errors (2)**
- Strategies to reduce patient specimen errors (0)**
- Practices to reduce patient identification errors (0)**
- Strategies to reduce patient identification errors (0)**
- Practices to reduce specimen identification errors (0)**
- Strategies to reduce specimen identification errors (0)**

- Reducing patient identification errors (16)***
- Reducing specimen identification errors (2)**
- Information systems AND laboratories AND reduce identification errors (0)**

*date/language exclusion criteria applied

**duplicate records removed

***both date/language exclusion and duplicate records removed

Exclusion Criteria

- Upon review of the title/abstract, exclude the article only if it can be determined that:
    - No practice was assessed (i.e., no outcome measures are identified)
    - The practice is not sufficiently described
    - The article is a commentary or opinion piece

Note to reviewers: It may not be apparent from the title and abstract if the exclusion criteria apply to the article.

Reviews: Reviewers will screen full-text references and apply exclusion criteria to identify those to be included in full-text reviews.

Reviewers will meet for consensus on references to be entered into the data abstraction database.

Inclusion Criteria

- The specific intervention/ practice identified in the literature:
    - Is in use and available for application
    - Can be performed and reproduced in other comparable patient care settings
    - Impacts a defined group of patients
    - Identifies a potential improvement in an outcome that can be related to at least one of the following aspects of patient care: effectiveness, efficiency, patient-centeredness, safety, timeliness or equity

Data Abstraction: References identified for abstraction will be entered into the data abstraction database.

Topic Area: Communicating Critical Values

Purpose: To identify peer-reviewed publications that assesses candidate practices for communication of critical values for laboratory tests.

Review Question: What practices are effective for communicating laboratory critical value results to the licensed caregiver who can act on them?

Search Strategy: A comprehensive search of English language literature published during or after 1995 using multiple databases and other strategies. These include:

a. PubMed, MedLine, CINAHL, and Cochrane databases

b. Professional guidelines electronic databases (CLSI, ISO, NACB)

c. Hand searching journals of relevance to the review topic, reports, conference proceedings, and technical reports.

d. Reference lists of relevant reviews

e. Key informants: consultation with sub workgroup for relevant information sources

The title/abstract of the article should address:

a. A critical value reporting practice,

b. Reporting or Communication of laboratory critical values/ critical test results,

c. Technology to improve critical values reporting/improve processes in result reporting,

d. Assessment/Monitoring of critical values reporting.

At least two reviewers will independently screen the search results by applying initial inclusion/exclusion criteria.

A pre-abstraction reference list of literature meeting the initial inclusion criteria will be generated indicating references for full-text review.

Initial Search Terms

- Communication of critical values in laboratory tests (29)
- Communicating critical values in laboratory tests (0)*
- Communicating critical values (12)***
- Critical value reporting (95)***
- Critical value identification (364)***
- Critical value identification AND reporting (0)**

* date/language exclusion criteria applied

**duplicate records removed

***both date/language exclusion and duplicate records removed

Initial Inclusion/Exclusion Criteria

Initial inclusion criteria

- The title/abstract of the article addresses:
    - A critical value reporting practice
    - Reporting or Communication of laboratory critical values/ critical test results
    - Technology to improve critical values reporting/improve processes in result reporting
    - Assessment/Monitoring of critical values reporting

Inclusion Criteria:

- At the minimum, the published information source should be an observational or descriptive study that describes pre-analytic practices of reporting and/or communicating critical values from laboratory tests to patient care setting.
- The specific intervention/ practice identified in the literature:
    - Is in use and available for application
    - Can be performed and reproduced in other comparable patient care settings
    - Impacts a defined group of patients
    - Identifies a potential improvement in an outcome that can be related to at least one of the following aspects of patient care: effectiveness, efficiency, patient-centeredness, safety, timeliness or equity

Initial exclusion criteria

- Upon review of the title/abstract, exclude the article only if it can be determined that:
    - No practice was assessed (i.e., no outcome measures are identified)
    - The practice is not sufficiently described
    - The article is a commentary or opinion piece

Note to reviewers: It may not be apparent from the title and abstract if the exclusion criteria apply to the article.

Reviews: Reviewers will screen full-text references and apply inclusion/exclusion criteria to identify those to be included in the data abstraction database.

Reviewers will meet for consensus on references to be entered into the data abstraction database.

Inclusion/Exclusion Criteria

References are to be excluded IF:

- The reference is a commentary/opinion piece about critical values reporting
- A post-analytic practice is not specified
- The practice is not sufficiently described (service, materials or policy implemented not described, how delivered/implemented not described)
- No practice assessed (no outcome measure is reported)

Data Abstraction: References identified for abstraction will be entered into the data abstraction database.

Topic Area: Blood Culture Contamination

Purpose: to identify candidate interventions/practices that are effective at reducing blood culture contamination.

Review Question: What interventions/practices are effective at reducing blood culture contamination?

Review Search Strategy: Focus on identifying relevant **published** and **unpublished** literature that categorizes/defines blood culture contamination and/or potential interventions/practices that reduce blood culture contamination. Information sources are to screened by the CDC Review Team to select articles/other literature for full-text review.

A comprehensive search of English language literature published after 1995 to present using multiple databases and other strategies. These include:

a. PubMed and Cochrane databases that often linked to related articles
b. Professional guidelines electronic databases (Cumitech, CLSI, ISO, NACB)
c. Hand searching journals of relevance to the review topic, reports, conference proceedings, and technical reports.
d. Reference lists of relevant reviews
e. Key informants: consultation with sub workgroup for relevant information sources

The title/abstract of **published** articles should address:

a. A blood culture contamination reduction practice,

b. Reporting of blood culture contamination test results,

c. Blood culture collection techniques aimed at reducing the blood culture contamination rates,

d. Assessment/Monitoring of patient infection rate due to collection site/technique.

Initial Search Terms

- Blood culture contamination (1677)
- Blood culture contamination rates(117)
- Blood culture contamination laboratory tests results (9)*
- Blood culture contamination indicators(50)***
- Causes of blood culture contamination (0)***
- Strategies to reduce blood culture contamination (15)***
- Proper site preparation procedure for blood culture collection (0)***

* date/language exclusion criteria applied

**duplicate records removed

***both date/language exclusion and duplicate records removed

Exclusion Criteria

Upon review of the title/abstract, exclude the article if:

- The title is not applicable to blood culture contamination

- No practice was assessed (i.e., no outcome measures are identified)

- The practice is not sufficiently described (service, materials or policy implemented not described, how delivered/implemented not described)

- The article is a commentary or opinion piece that contains no specific information on costs, benefits, and/or implementation.

Reviews: Reviewers will screen full-text articles, apply exclusion criteria and meet for consensus on references to be entered into the data abstraction database.

Inclusion Criteria

- At the minimum, the published information source should be an observational or descriptive study
- The specific intervention/ practice identified in the literature:
    - Is in use and available for application
    - Can be performed and reproduced in other comparable patient care settings
    - Impacts a defined group of patients
    - Identifies a potential improvement in an outcome that can be related to at least one of the following aspects of patient care: effectiveness, efficiency, patient-centeredness, safety, timeliness or equity

APPENDIX E. EVIDENCE CONSENSUS RATINGS AND SUMMARY TABLES

This page intentionally left blank

CONTINUE ON NEXT PAGE

Practices that are effective for reducing PATIENT SPECIMEN IDENTIFICATION errors

Laboratory Medicine Best Practices

Consensus Ratings 2009

TOPIC AREA: Patient Specimen Identification

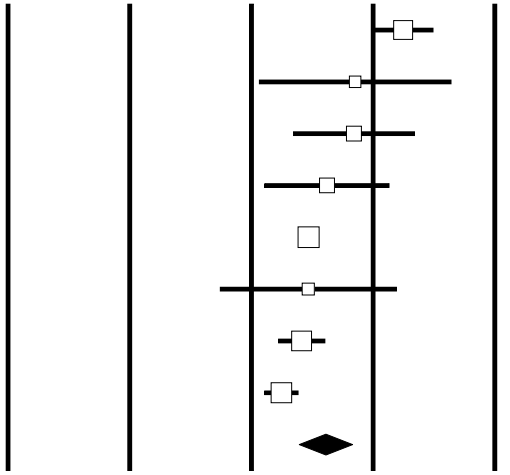
Practice: Bar Coding Systems	Study Quality Rating						Effect Size Rating	Overall Consistency	Overall Strength of Body of Evidence
	Study	Practice	Measures	Results	Total	Rating			
Bologna 2002	2	2	2	2	8	Good	Substantial	Yes	5 Studies = Good/Substantial 1 Study = Good/Moderate 2 Studies = Fair/Substantial 2 Studies = Poor/na High
Hayden et al. 2008	3	2	2	3	10	Good	Substantial		
Killeen et al. 2005	2	2	2	3	9	Good	Substantial		
Sandler et al. 2005	1	1	2	0	4	Poor	n/a data insufficient		
Turner et al. 2003	1	1	1	1	4	Poor	n/a		
Zarbo et al. 2009	2	2	2	3	9	Good	Moderate		
Unpub A 2009	3	1	1	2	7	Fair	Substantial		
U of MN 2009	1	2	1	1	5	Fair	Substantial		
U of WA	2	2	2	2	8	Good	Substantial		
LBJ 2009	2	2	2	2	8	Good	Substantial		

Patient Specimen Identification: Bar Coding Systems

Model	Study name	Log odds ratio	Lower limit	Upper limit
	U WA 2009 *	4.99	4.01	5.97
	U MN 2009 *+	3.41	0.26	6.56
	Bologna 2002 **	3.37	1.38	5.36
	Unpublished PSID A *	2.49	0.44	4.53
	Hayden 2008 **	1.88	1.66	2.11
	LBJ General 2009 *	1.87	-1.02	4.77
	Killeen 2005 **	1.65	0.89	2.41
	Zarbo 2009 **	0.99	0.44	1.53
Random		2.45	1.57	3.33

Statistics for each study

Log odds ratio and 95% CI



**="Good" Study Quality rating
*="Fair" Quality rating
+=Variance Uncertain

The sample size is reflected by the size of the rectangle; uncertainty (95% confidence interval) is represented by horizontal lines.
The diamond is the overall effect size from pooling of the evidence. Boxes are proportional to weights.

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Bologna LJ (1) and Mutter M (1) - 2002 - J Healthcare Information Manag. - [1] The Valley Hospital, Ridgewood NJ **and** Bologna LJ (1), Lind C (1), and Riggs RC (2) - 2002 - Clin Leadersh Manag Rev. - [1] The Valley Hospital, Ridgewood NJ; [2] BD Strategies for Becton Dickinson, Franklin Lakes, NJ - Funding not reported	- **Design:** Before-After - **Facility/Setting**: Valley Hospital, Ridgewood, NJ, >400 bed community hospital - **Time period:** 7/97-7/98 and 1/2000-9/2000 (non-consecutive periods) **- Population/Sample:** Number of phlebotomies for 10 care centers: Before practice (1999): 69,432 After practice (1– 9/2000): 59,490 - **Comparator:** Combination of print and hand-written labels used **- Study bias:** none	**-Description:** Barcoding system; labels print at bedside; - **Duration:** 10 months (12/1999 – 9/2000) - **Training:** not discussed - **Staff/Other resources:** Nursing supervisor, licensed medical practitioners, phlebotomists; portable label printers, BD System Software; bar-coded wristbands - **Cost:** not reported	**- Outcome Measures:** Error Rates (1) Incorrect/incomplete label (2) Misidentified patients (3) Unnecessary phlebotomy - **Recording Method:** Internal Quality Control instrument	- Pretest-Posttest - **Findings/Effect Size:** Error Rates for the 10 care centers: (1) Incorrect/ incomplete label: 59% reduction pretest: 0.017% (12/69,432) posttest: 0.003% (1/59,490) ➢ **OR = 5.75** (2) Misidentified patients: 94% reduction, pretest: 0.049% (34/69,432) posttest: 0.003% (1/59,490) ➢ **OR = 29 (CI, 4 – 212)** (3) Unnecessary phlebotomy: 89% reduction pretest: 0.027% (19/69,432) posttest: 0.003% (2/59,490) ➢ **OR = 16.01** - **Statistical Significance/Test(s):** not discussed **- Results/conclusion biases:** None
Quality Rating: 8 (Good) **(10 point maximum)** **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2;** - Data collected during notably different time periods	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts. maximum): 2**	**Results/findings (3 pts maximum): 2;** - No statistical test results provided

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Hayden RT, Patterson DJ, Jay DW, Cross C, Dotson P, Possel RE, Srivastava DK, Mirro J, and Shenep JL. - 2008 - *Journal of Pediatrics* - [St. Jude's Children's Research Hospital (multiple departments), Memphis, TN, USA - Funding: Partly self-financed; and supported by the American Lebanese Syrian Associated Charities	- **Design:** Before-after - **Facility/Setting**: St. Jude Children's Research Hospital, pediatric cancer center with a high percentage of acutely ill patients; Memphis, TN - **Time period:** 36 months (9/03-8/06) - **Sample:** Pre-practice: 19,247 accessions (test order events) per month for 12 mos. (9/03 – 8/04) Post-practice: 17,793 accessions per month for 12 mos. (9/05 -8/06) - **Comparator**: Not reported - **Study bias:** none	- **Description:** Barcoding system using handheld data terminals (personal digital assistants, PDAs) with built in scanner that track and verify orders at the point of collection located in each inpatient room, each clinic and procedure room. Full implementation – 8/05. - **Duration:** 24 months - **Training:** Education provided using the "train-the-trainer" approach - **Staff/Other Resources**: Nurses. - **Cost:** Not reported	- **Outcome Measures:** PSID Error Rate: Mislabeled samples (%) - mismatches between patient name and. specimen (wrong label or specimen collected from wrong patient). - **Recording Methods:** Internal quality control instrument	- Pretest-Posttest - **Findings/Effect Size:** 84% reduction in PSID monthly mean error rate: Pre: 0.032% Post: 0.005% ➢ **OR = 6.58 (CI, 5.26 – 8.22)** - **Statistical Significance/Test(s):** P<.001 Nonparametric test, Post-hoc analysis with Wilcoxon rank sum test analysis. - **Results/conclusion biases:** Conclusions focus mostly on findings favorable to the practice described.
Quality Rating: 9 (Good) (10 point maximum) **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2;** -No comparator information provided	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Killeen JP, Chan TC, Jones K, and Guess DA - 2005 - *Academy of Emergency Medicine* - University of California, San Diego, CA, USA - Funding: Self-financed	- **Design:** Before-after - **Facility/Setting**: UCSD Medical Center, San Diego CA –**Time Period:** Two 6-month periods: 6 mos. pre-practice and 6 mos. post-practice (dates not reported) - **Sample:** All Emergency Department (ED) patients seen during study period (annual census: 40,000) Total ED laboratory specimens: Pre-practice: 22,243 Post-practice: 22,574 - **Comparator:** Imprint stamp sticker labels on specimens and paper requisitions **-Study bias:** Unclear if the 6-month periods being compared are immediately before and after implementation.	- **Description:** Barcoding system - **Duration:** 6 months duration; no dates provided -**Training:** not discussed **- Staff/Other Resources:** not discussed - **Cost:** not reported	**- Outcome Measure:** PSID Error Rate - Number of misidentified, unlabeled, or mislabeled specimens / number of specimens (per 1000 specimens) - **Recording Methods:** Occurrence log	- Pretest-Posttest - **Findings/Effect Size:** Pretest : 2.56 per 1000 [CI: 1.94 to 3.32] (0.00256%); 57/22,243; ID errors: 41 mislabeled and 16 unlabeled Posttest: 0.49 per 1000 [CI: 0.24 to 0.87], (0.00049%); 11/22,574; ID errors: 8 mislabeled and 3 unlabeled. - **Statistical Significance/ Test(s)**: p <.05 ; Chi-square test ➢ **OR = 5.21 (CI, 2.44 – 11.11)** - **Results/conclusion biases:** None reported
Quality Rating: 9 (Good) (10 point maximum) **Effect Size Magnitude Rating: Substantial (Relevance: Less Direct)**	**Study (3 pts maximum): 2;** - Study/practice dates not provided	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Sandler SG, Langeberg A, and Dohnalek L. - 2005 - *Developmental Biology* - Georgetown University Hospital (multiple departments), Washington DC, USA - **Funding:** Partially from the Greenspring Financial Insurance Limited (GFIL)	- **Design:** Non-comparative study - **Facility:** Georgetown University Hospital, Northwest Washington, DC; 609-licensed bed, not-for-profit, acute care teaching and research facility. - **Study Setting**: "18-bed hematology-oncology-bone marrow transplant unit staffed by approximately 20 licensed registered nurses". - **Time Period:** 10/02- no end date provided - **Sample**: 125 tests, all blood samples and blood components for transfusions. -**Comparator:** Not reported **- Study bias:** Small sample size, no comparison data or complete time period provided. The number of patients represented by transfusions is not reported.	- **Description:** Barcoding system for transfusion linking patients' wristbands with blood component labels. Consists of the hand-held PC/bar-code scanner with radio frequency port to a portable printer. Software checks for: (i) the operator's electronic signature (personal ID badge bar code); (ii) the patient's name and medical record # (wristband); (iii) the blood component (compatibility bar code); and (iv) the blood centre's whole blood # (blood bag bar code). - **Duration:** 10/02 - ?(no end date) - **Training:** provided during 1-hour session including written and instruction review on how to use the system. - **Staff:** Nurses - **Cost:** not reported	**- Outcome Measures:** (1) Positive Identification rate Percentage of patients, blood samples and blood components for transfusion positively and accurately identified (2) Number of correctly labeled samples - labels for blood sample tubes & certification forms legible with complete information **- Recording Method:** electronic medical record	- Non-comparative Study, Time series (average): - **Findings/Effect Size:** (1) "All (100%) patients, blood samples, and blood components for transfusion were positively and accurately identified." (2) "All (100%) bar-code-labeled blood sample tubes and certification forms were legible with complete information. - **Stat. Significance/Test(s)**: None - **Results/conclusions biases**: The stated purpose was to focus on nurses who transfuse blood infrequently, yet no statistics presented for these results (suggest that these nurses perform more poorly than nurses who transfuse frequently). Results focused on subjective ratings.
Quality Rating: 3 (Poor) (10 point maximum) **Effect Size Magnitude Rating: N/A (data insufficient)** **(Relevance: Direct)**	**Study (3 pts maximum): 1;** - Complete study time period not reported - Transfusion study may be too distinctive to be generalizable	**Practice (2 pts maximum): 1;** No practice duration specified	**Outcome measures (2 pts maximum): 1;** Recording method is not adequately described.	**Results/findings (3 pts maximum): 0** - Insufficient sample: Statistical power not discussed and sample size too small - Data insufficient to allow effect size calculation(non comparative study)

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Turner CL, Casband AC, Murphy MF [1] - Yr Published: 2003 - Publication: *Transfusion* - [1] National Blood Service, John Radcliffe Hospital Oxford, UK **- Funding:** National Blood Service	- **Design:** Observational study - **Facility/Setting** Oxford Radcliffe Hospital, 1500 bed teaching hospital, Oxford, UK; Setting: Hematology outpatient clinic (later extended to hematology inpatient ward) - **Study Period:** Not reported - **Sample:** First unit RBC transfusions: Pre: 51 (48 blood prescribed) Post: 51. (45 blood prescribed) Sample collection: Pre: 30; Post: 30 - **Comparator:** Standard system without barcoding (checking and administering blood total process of 27 defined steps; sample collection process 17 steps); manual checking/verification of patient information on patient wristband and patient chart	- Barcoding system using hand-held computers for scanning pdf barcodes, generates barcoded wristbands and labels via portable printer for crossmatch with blood administration and sample collection process verification steps at the patient bedside. Checking and administering blood total process of 16 defined steps; sample collection process 8 steps. - **Duration:** Not provided - **Training:** Education/training on transfusion safety and use of the barcode system was provided to staff - **Staff:** Phlebotomists, blood bank receptionists, IT, the blood bank (Note: Staff preferred the new technology once familiar with it) - **Cost:** Initial equipment/support ~ $1.2 million (US$ 2003)	**- Outcome Measures:** Blood administration (pre-transfusion) – Percent correct performance of blood pack bedside check: (1) Patient's ID (surname, first name, date of birth, sex, and hospital number). (2) Cross ref. of blood group, unit #,compatibility label, expiration date, patient's prescription & transfusion report form special requirements (3) Sample collection: Percent sample tubes labeled immediately with hospital number, surname, first name, date of birth, sex, sample date **- Recording Method:** Audits/direct observations	- Pretest-Posttest: - **Findings/Effect Size:** (1) Blood admin. patient ID check: Pre: 100% (51 /51) Post: 100% (51 /51) 0% improvement (2) Blood admin. cross reference check: Pre: 9.8% (5/51) Post: 41.2% (21 /51) 30.4% improvement p-value : 0.0005 (Table 2, p. 1205) (3) Sample collection labels - patient ID -Pre: 50% (15/30) Post: 100% (30/30) 50% improvement p-value: <0.0001 (Table 4, p. 1206) - **Stat. Significance/Test(s):** Stat. analysis using exact tests of independent proportions -**Biases:** Study period not reported, small sample, barcoding system reinforced with education and training
Quality Rating: 4 (Poor) (10 point maximum) **Effect Size Magnitude Rating: N/A** **(Relevance: Indirect)**	**Study (3 pts maximum): 1;** - Transfusion study may be too distinctive to be generalizable - Study design, time period and sample selection methods may introduce bias affecting results	**Practice (2 pts maximum): 1;** - Important practice/implementation characteristic (dates/duration) not identified	**Outcome measures (2 pts maximum): 1;** - Process compliance outcome measures only modestly related to evidence review question	**Results/findings (3 pts maximum): 1;** - Potentially insufficient sample: measurement period; # of subjects not reported; sample size - Staff training may impact effect size; not clearly attributable to barcoding

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Zarbo RJ, Tuthill JM, D'Angelo R, Varney R, Mahar B, Neuman C, Ormsby A. - 2009 - *Am J Clin Pathol* - Department of Pathology, Henry Ford Hospital, Detroit, MI. - **Funding:** Not reported	- **Design:** Before-after - **Facility:** Henry Ford Hospital, Detroit, MI. - **Study Setting:** Surgical Pathology lab gross room - **Time Period:** Two 3-week periods: Pre: 7/06; Post: 8/07 - **Sample:** Cases: Pre: 2,694; Post: 2,877 Specimen parts: Pre: 4,413; Post: 4,725 Tissue cassettes: Pre: 8,776; Post: 9,167 Histology slides: Pre: 14,270; Post:17,927 - **Comparator:** Simple-logic, bar-coded slide label only (2006); specimen accessioning and processing completed by manual entry of information and hand written labels on specimen cassettes and slides.	- Barcoding system and process redesign to standardize workflow using a complex-logic, bar-coded pathway tying together 4 work cells to provide computer-readable encoding for identification of parts, and accession, gross dissection, histology/microtomy, and pathology sign-out stations. Also includes manual quality control checks at each station to ensure case integrity before processing, and barcoding in all subsequent processes (2007 implementation) - **Duration:** 3 weeks - July 2007 - **Training:** Group education session, ensuring all staff members were in unison on the goals and time frame of the data collection and how to use the visual data display - **Staff:** Surgical pathology, histology and informatics staff - **Cost:** Not reported	**- Outcome Measures:** Patient specimen identification (PSID) error rates (1) Surgical cases (2) Specimen parts – mismatch between pathology requisition and patient information (3) tissue cassettes – mismatch between cassette ID and lab tag information (4) histology slide labels - **Recording Method:** Data collected, recorded and defects categorized by 59 surgical pathology personnel (21 senior staff and 38 technical staff), using a visual data display collection tool (details on page 469)	- Pretest-Posttest - **Findings/Effect Size:** PSID error rates(1) Surgical cases: Pre (2006): 1.67% (45/2,694) Post: (2007):0.63% (18/2,877) 62.3% reduction; p-value: <.001 ➢ **OR = 2.68 (CI, 1.55 – 4.63)** (2) Specimen parts Pre (2006): 0.23% (10 /4,413) Post (2007): 0.38%, (18 /4,725) Not statistically significant (3) Tissue cassettes Pre (2006): 0.057% (5 errors/8,776) Post (2007): 0.055% (5 errors/9,167) 3.5% reduction; Not statistically sign. (4) Histology slide labels Pre(2006): 0.21%, (30 errors/14,270) Post (2007): 0.01% (2 errors/17,927) 95% reduction; p-value <.001 - **Stat. Significance/Test(s):** χ^2 tests (Fisher exact test adjusted for small counts and Mantel-Haenszel test) to 2 data sets
Quality Rating: 9 (Good) **(10 point maximum)** **Effect Size Magnitude Rating: Moderate** **(Relevance: Direct)**	**Study (3 pts maximum): 2;** - Workflow process (surgical pathology) may be too distinctive to be generalizable to other barcoding practices	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Unpublished Study A – Barcoding System - 2009 - LMBP Network - In-house audit or quality control initiative - Western USA - **Funding:** Self-financed	- **Design:** Before-after - **Facility:** Academic medical center in Western U.S.; >300 beds; > 1,000,000 lab tests/yr. **- Setting:** After 1 mo. med./surg. unit pilot (11-12//2006) implemented on other non-ICU units with laboratory phlebotomy team - **Time Period:** 1/1/06 -7/31/2009 - **Sample:** All patient blood specimens (sample size not reported); approx. 33% of 29,000/mo. collected by phlebotomy team (units served: neurology, med./ surg., transplant, ob/gyn, oncology, neuropsych, emergency, pediatric, and nicu) Estimated sample: 9,570/mo. (.0.33 x 29,000); 114,840/year - **Comparator:** Status Quo (no barcode labeling system)	- **Description:** Barcoding system consists of phlebotomists using patient bedside barcode specimen labeling with wireless handheld device and an attached mini-barcode label printer. The device can access the patient test orders in real time, collect orders, and print test labels at the patient bedside. - **Duration:** 11/06-7/09, ongoing -**Training:** 3 trainers provided directly from system vendor; no details on person-time or intensity. - **Staff:** 2 FTEs from clinical lab IT; 1 FTE phlebotomy supervisor; 20 FTEs around-the-clock (24/7) blood draws - **Cost:** Start-up software: $30,000; hardware: $72,000; Annual maintenance: $32,000 (US$ 2006)	**- Outcome Measure:** 1) Annual # of patient specimen Identification (PSID)errors Note: error rate estimated from data provided by authors: - **Recording Methods:** Event reporting system and occurrence management reports log	- Pretest-Posttest **- Findings/Effect Size:** Total (annual) PSID errors reported Pre: 2006: 12 errors Post: 2007: 1 error 2008: 0 errors; 2009: 0 errors (through 7/09). PSID Error Rate (calculated using above PSID errors using estimated sample size of 114,840/ yr.): Pre-: 12/114,840 = 0.010% Post-: 1/114,840 = 0.0008% **OR = 12.00 (CI,1.56 – 92.3)** - **Stat. Significance/ Test(s):** None reported **Results/conclusion bias:** None reported
Quality Rating: 7 (Fair) **(10 point maximum)** **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 3**	**Practice (2 pts maximum): 1;** - Description lacks detailed specifications on how system interfaces with hospital & requirements for implementation	**Outcome measures (2 pts maximum): 1;** **-** Recording method may not accurately capture all instances of the outcome	**Results/findings (3 pts maximum): 2;** - Small number of errors reported yields unstable effect size estimate

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Unpublished Univ. of Minnesota Medical Center Fairview, Acute Care Laboratory, Minneapolis, MN, USA - 2009 - LMBP Network - In-house audit or quality control initiative - **Funding:** Self-financed	- **Design:** Case-Control - **Facility/ Setting:** UMN Fairview, Minneapolis, MN; >300 bed academic medical center; > 1,000,000 tests/yr. **- Study Setting**: Clinical lab, ED, Adult and Pediatric ICUs - **Study Period:** Multiple study arms/start dates in various units: Clinical lab: 06/2006 -08/2009; ED 07/2006- present; PCU pilot only: 11/2006-2/2007; Adult ICUs: 02/2008-08/2009; No end date reported. **- Sample:** Includes 100% of specimen (container)s and all error types. Specimen types, numbers and dates not reported. Total volume: 39,300 / month. **- Comparator:** No details reported for practice - **Study bias**: Time period and sample selection methods may introduce bias affecting results	- **Description:** Barcoding system for patient id using hand-held PC to verify specimen labels match wristband prior to labeling specimen tubes at the bedside - **Duration:** 3 years, first implemented in the clinical lab (06/2006- present, ongoing), then campus ED (07/2006-present), PCU pilot only (11/2006-2/2007) Adult ICUs (02/2008-08/2009); practice is ongoing. - **Training:** ~30 min. for users - **Staff:** Lab in collaboration with nursing and IT staff to implement 0.5 FTE maintaining, auditing, problem-solving, validation, etc. - **Cost:** Start-Up: Design & programming cost: $600,000; hardware : ~ $425,000 Post Start-Up: ~$425,000 for new installations; ~$300,000 for replacement hardware	**- Outcome Measure:** Patient specimen Identification (PSID) error rate: Number of mislabeled specimens, wrong specimen in tube (WSIT)and unlabeled specimens per 10,000 collections - **Recording Methods:** Pre-implementation: manual error reporting system Post-implementation: electronic reporting system (electronic event tracking logs, and compared to "cancel comments" in lab computer system). Compliance (scan rate) based on monthly 1-day audits of each unit where barcoding is implemented.	- Pretest-Posttest **- Findings/Effect Size** PSID error rate Units Without Barcoding system: 12.1 errors/10,000 collections. Units with Barcoding system: 0.4 errors/10,000 collections - **Stat. Significance/Test(s):** Proportion successful; significance not reported **Results/conclusion biases:** Sample sizes and specific dates not reported. Results not specified by medical unit (i.e., with/without barcoding system vs. those not reported). WSIT and unlabeled specimens in numerator (outcome measure); not relevant to review question
Quality Rating: 5 (Fair) **(10 point maximum)** **Effect Size Magnitude Rating: Substantial** **(Relevance: Less Direct)**	**Study (3 pts maximum): 1;** - Sample not adequately described; may not represent results of the practice	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 1;** - Recording method change pre vs. post could impact error results	**Results/findings (3 pts maximum): 1;** - Measurement period not clearly defined, results not specified; results may not be due to practice - Statistical power not discussed

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Unpublished Univ. of Washington, Pathology Laboratory, Seattle, WA, USA - 2009 - LMBP Network - In-house audit or quality control initiative - **Funding:** Self-financed	- **Design:** Before-after - **Facility:** Univ. of Washington, Seattle, WA;,>300 bed academic medical center; > 1,000,000 lab tests annually -**Study Setting:** Anatomic pathology lab gross room. - **Time Period:** 12/1/2007 – 09/2009 - **Sample:** Patient specimen cassettes/blocks (includes outpatient biopsies and inpatient surgical specimens). Pre: 12/2007– 2/2008; 85,213 Post: 12/2008- 9/2009, 50,016 Historical annual baseline volume: 85,213 cassettes/blocks produced - **Comparator:** Not described - **Study bias:** Findings for personnel savings based on user recall	-Barcoding system to identify and track all specimens from accessioning to gross station, and just-in-time, single piece workflow system for cassette/block labeling, with computer printing 2D barcoded cassettes at grossing station requiring custom software and commercial cassette printers. - **Duration:** 9 months (12/1/08-9/09), ongoing; initial partial implementation in 12/2008, full implementation 3/2009. -**Training:** End User Training for permanent gross room personnel and rotating pathology residents - **Staff:** Gross room pathologists, path. assistants and residents - **Cost:** Software custom ~$200,000; 4 cassette printers @ $20,000 each = $80,000; Hardware: PCs, barcode readers, label printers, mounting arms = $8,000 (US$ 2008)	**- Outcome Measures:** (1) Patient specimen identification (PSID) error rate: Number mislabeled specimen cassettes(includes duplicate number, wrong specimen, wrong case, wrong patient) / total number of pathology specimen cassettes (2) Personnel Savings – estimate of how much labor hours saved due to implemented practice - **Recording Methods:** (1)Incident Reports (pre and post); also counted directly post only (2) Survey of gross room personnel – estimates of time saved due to barcoding system	- Pretest-Posttest - **Findings/Effect Size:** (1) PSID error rate Pre Barcode System (12/07 -12/08): 1.16% (988/85,213) Post Barcode System (12/0 – 09/09): 0.00080% (4/50,016) ➢ **OR = 147 (CI, 55 – 391)** (2) Post-practice: saved 0.75-1.0 FTE gross room personnel (less material handling, less error resolution efforts), not reported over what time period. - **Stat. Significance/Test(s):** Proportion successful; no statistical analysis/significance reported - **Results/conclusion bias:** personnel savings based on user recall (no point deduction as bias is for non-effectiveness measure – feasibility)
Quality Rating: 8 (Good) **(10 point maximum)** **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2;** - Workflow process (anatomic pathology) may be too distinctive to be generalizable to other barcoding practices	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 2;** - No statistical analysis/significance

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Unpublished Lyndon B Johnson General Hospital Core Laboratory, Houston, TX, USA - 2009 - LMBP Network - In-house audit or quality control - **Funding:** Self-financed	- **Design:** Before-after - **Facility:** Lyndon B Johnson hospital, Houston, TX, >300 bed teaching hospital.; > 1,000,000 lab tests annually **Study Setting**: All nursing units except ER, ICU, NICU and Outpatients; Lab collects 8,500 specimens monthly, 280/ daily - **Time Period:** 1/1/2009 – 8/31/2009 Pre: :1/2009 – 4/2009 (4 mos.) Post: 6/2009 –8/2009 (3 mos.) - **Sample:** Inpatient blood samples by venipuncture Pre: 41,815 Post: 24,789 - **Comparator:** Phlebotomists use of a printed draw list and pre-printed specimen label to enter collection information (date/time & ID of patient) - **Study bias:** Low baseline error rates may underestimate effectiveness in general use.	- Barcoding system used by laboratory department only; laboratory phlebotomists print labels from wireless handheld printer and label specimen tubes by the patients' bedside; in use 24/7. - **Duration:** 6/1/2009- 10/1/2009 - **Training:** Staff training takes 3 hours to learn equipment use - **Staff:** 20 phlebotomists, IT facility staff for installs and training - **Cost:** Cost related to training phlebotomists: $14.20 * 3 hours * 20 FTEs =$852. Cost of Collection Manager (hardware, installation, support, and training) = $1 million for district (2 hospitals; 650 and 330 beds respectively). (US$ 2009)	**- Outcome Measures:** Patient specimen identification (PSID) error rate: Number of mislabeled specimens/total number of specimens - **Recording Methods:** Incident reports Pre: review of occurrence log based on manual forms Post: online application	- Pretest-Posttest **Findings/Effect Size:** PSID error rate Pre: 1.012%, (5/ 41, 815) Post : 0.00% (0/24,789) 100% positive identification; **OR = 6.50 (CI,0.36 – 117.61)** **Stat. Significance/Test(s):** Proportion successful/ significance not reported **Results/conclusion bias:** None reported
Quality Rating: 8 (Good) **(10 point maximum)** **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2;** - Study bias – phlebotomists only with low initial error rates	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 2;** - Statistical analysis/power not reported

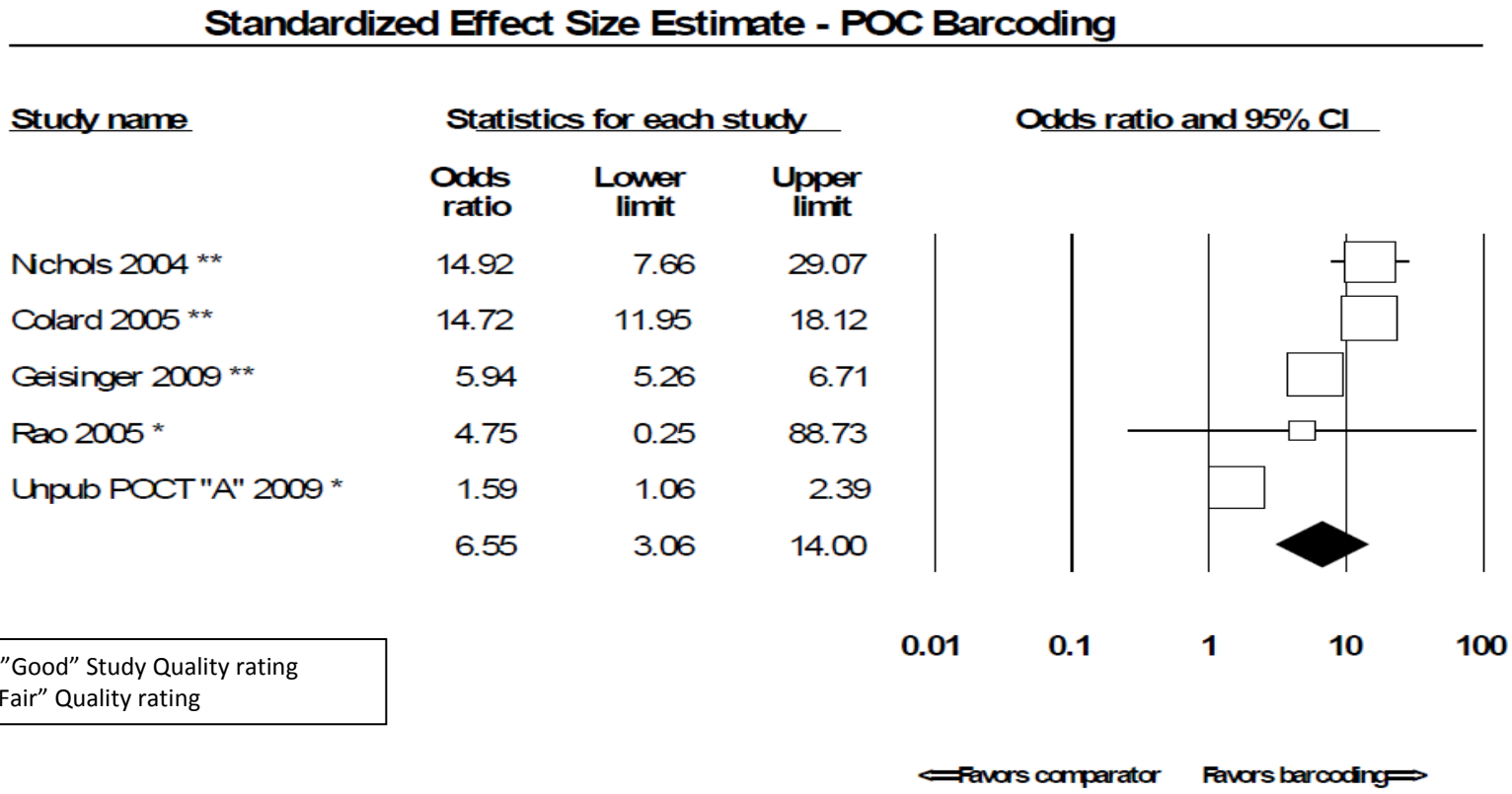
Laboratory Medicine Best Practices

Consensus Ratings 2009

POINT OF CARE TESTING BAR-CODING SYSTEMS

Practice: POCT Bar Coding Systems	Study Quality Rating						Effect Size Rating	Overall Consistency	Overall Strength of Body of Evidence
	Study	Practice	Measures	Results	Total	Rating			
Colard 2005	2	2	2	2	8	Good	Substantial	Yes	3 Studies = Good/Substantial 1 Study = Fair/Substantial 1 Study = Fair/Moderate High
Nichols et.al 2004	3	2	1	2	8	Good	Substantial		
Rao et al. 2005	1	2	1	2	6	Fair	Substantial		
Geisinger 2009	2	2	2	2	8	Good	Substantial		
Unpub B 2009	2	2	2	1	7	Fair	Moderate		

Patient Specimen Identification: Point-of-Care-Testing Bar Coding Systems



**="Good" Study Quality rating
*="Fair" Quality rating

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Sample - Comparator - Study Bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Colard, DR [1] - 2005 - Point of Care [1] Department of Pathology, Saint Luke's Hospital; Kansas City, MO - Funding: not reported	- Design: Before-after - Facility/Setting: Saint Luke's Hospital, Kansas City, MO; 629-bed tertiary care teaching hospital for the Univ. of Missouri-Kansas City School of Medicine. - Time Period: 10/2000 - 5/2004 Pre: 10/2000 – 11/2002 (25 mos.) Post: 12/2002 – 5/2004 (17 mos.) - Sample: The point-of-care glucose program at St. Luke's Hospital, Kansas City, MO, performs 12,000 tests per month. All glucose POCT taken within given time period. No data provided on number of tests or tests per month for study sample. - Comparator: Manual entry of patient information and test results into the laboratory information system. - Study Bias: unexplained sample size increase; gap in outcome data ((12/2003 – 5/2004)	- Description: POCT Barcoding - Duration: 19 months (12/2002-5/2004) - Training: training provided to all nursing staff; nurses with high scan errors received additional training. Due to modifications post-implementation, additional training required. - Staff/Other Resources: nursing staff (operators); POCT coordinator implement new process change - Cost: not reported	- Outcome measures: (1) Patient ID Error Rate - % patient identification error rate for point of care blood glucose tests (2) Number unidentified point of care blood glucose tests - Recording method: Occurrence log	- Pretest-Posttest (1) Monthly error rates- Pre-barcoding: 9.4% Post-barcoding: 0.7%; ➢ OR = 14.72 (CI, 11.95 – 18.12) (2) Monthly unidentified test counts – Pre-barcoding (11/2002): 404 Post-barcoding (5/2004): 6 ➢ OR = 69.6444; d = 2.340) - Stat. Significance/Test(s): Not discussed - Results/Conclusion Bias: -Data presented as reported above with corresponding monthly "process changes" - Sample size not explicitly reported; effect size calculated based on estimate from authors' data.
Quality Rating (10 point maximum): 8 (Good) **Effect Size Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2**	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 2;** **- Appropriateness of statistical analysis;**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Sample - Comparator - Study Bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Nichols JH [1,2]; Bartholomew C [2]; Brunton M [2]; Cintron C [2] ; Elliott S [2], et al - 2004 - Clinical Leadership Management Review [1] Tufts University School of Medicine. [2] Baystate Health System in Springfield, MA. - Funding: Self-financed	- Design: Before-After - Description: Baystate Health System, integrated delivery network, in western MA; 3 hospitals; > 850 beds; (Baystate Med. Ctr., Franklin Med. Ctr., Mary Lane Hospital); almost 1 million point-of-care tests/yr - Time Period: 1/ 2002 – 1/2004 Pre: 1/2002 – 10/2002 Post: 11/2002 – 1/2004 - Sample: Intensive Care Unit (ICU) blood gas and glucose POCTs, system wide volumes of nearly 600,000 glucose and blood gas tests annually for each year of the study period. All tests in sample (all 3 hospitals) - Comparator: Status quo (POCT without barcoding) and operator lock out "3-Strike Rule" - Study Bias: operators continued to have intermittent problems reading bar-coded wristbands.	- Description: POCT Barcoding 5-digit operator code and 9-digit patient account number - Duration: 15 months (11/ 02 – 1/04) - Training: Not reported - Staff/Other Resources: Nursing and pathology departments - Cost: Not reported	- Outcome measure: Patient ID errors : monthly number (count): (1) Glucose meter (2) Blood gas POC devices - Recording method: Internal quality control instrument	- Pretest-Posttest: - Findings/Effect Size: Identification errors (1) "Rates of identification errors decreased significantly over time for…glucose devices after implementation of bar coding (p = 0.0007)..."* decreased from average 26/month to 1/month (2) "Rates of identification errors decreased significantly over time for…blood gas...devices after implementation of bar coding (p = 0.048)*, from avg. of 4.6 to 1.7/month ➢ OR = 14.92 (CIO,7.66 – 29.07) - Stat. Significance/Test(s): Not reported; p-values provided - Biases: *Sample sizes or monthly volumes not provided. Unclear from the data presented whether compare bar coding to the lock-out program only (June-Oct. 02) or to cumulative data from both the lock-out program and previous period – Jan-Oct. 02.)
Quality Rating (10 point maximum): 8 (Good) **Effect Size Rating: Substantial (Relevance: Direct)**	**Study (3 pts maximum): 3**	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 1;** **- Face validity: Monthly errors without monthly test volume not error rates**	**Results/findings (3 pts maximum): 2; - Appropriateness of statistical analysis: Does not provide data sufficient to allow/verify calculation of an effect size (sample size not reported)**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Sample - Comparator - Study Bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Rao, AC; Burke, DA; and Dighe, AS [1] - 2005 - Point of Care [1] Massachusetts General Hospital - Funding: Self-financed	- Design: Before-after - Facility/Setting: Massachusetts General, Boston, MA; 900-beds; largest teaching hospital of Harvard Medical School and hospital-based research program in the U.S. - Time Period: No dates reported; 1 month Pretest 1 month Pilot test – bar coding 1 month Posttest - Sample: 35 inpatients included in the pilot test of bar coding only - 462 total glucometry tests: - Pre: 170 - Pilot: 134 - Post: 150 - Comparator: Usual Care - POC device with keypad for manual data entry for patient identification - Study Bias: none	- Description: POCT Barcoding 2D bar code for patient wristbands; only the medical record number included in the ID bar code which is a unique identifier for each patient. - Duration: 1 month (pilot test only) - Training: Provided to teach nurses, operations coordinators, operations assistants on how to print wristbands and troubleshoot. - Staff/Other Resources: Not reported. - Cost: Not reported	- Outcome Measure: Patient ID error rate - % errors in Medical Record Number (MRN) - Recording method: Verification of MRN; not described	- Pretest-Posttest - Findings/Effect Size: 1 month preceding test: 1.2% (2/170) 1 month pilot test: 1.5% (2/134) 1-month test: 0% (0/158) - Stat Significance/ Tests: Difference in error rates was statistically significant by Chi-squared analysis (P<0.005). ➢ OR = 4.75 (CI, 0.25 – 88.73) (results for 2 comparator periods pooled) - Results/Conclusions Biases: Study period is short, 1 month; relatively small sample sizes.
Quality Rating (10 point maximum): 6 (Fair) **Effect Size Magnitude Rating: Substantial (Relevance: Direct)**	**Study (3 pts maximum): 1;** **- The sample may not be representative of the results of the practice. The study time period and sample selection methods may introduce a study bias that could affect results**	**Practice (2 pts maximum): 2;**	**Outcome measures (2 pts maximum): 1;** **- Recording method not described**	**Results/findings (3 pts maximum): 2;** **- Measurement period insufficient and sample may be too small to allow a robust estimate of the impact of a practice**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Sample - Comparator - Study Bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Unpublished - Geisinger Medical Center: Schuerch, C [1] - 2009 - LMBP Network [1] Department of Laboratory Medicine, Geisinger Medical Center, Danville, PA - Funding: Self-funded	- Design: Observational study - Facility/Setting: Geisinger Medical Center, Danville, PA; a teaching hospital; > 300 beds; > 1 million tests annually. - Time Period: 1/2004 -6/2009 Baseline (initial barcoding implementation): 1/2004 (1 mo) Barcoding practice (full implementation): 1- 6/2009 (6 mos)* - Population/ Sample: Point-of-Care glucose tests for hospital inpatients. On average, approximately 18,000 inpatient POC glucose tests/month; total glucose tests during 6 months 2009: 106,780 - Comparator: Initial stage of barcode POCT implementation (2002-2004) compared to full practice implementation (2007-2009). Improvement Committee, investigation, reporting; monitoring after low scan rates (< 1/3) and high ID errors - Study Bias: Data used do not reflect pre-barcoding and based on 1 month following 2 years of implementation.	- Description: POCT Barcoding with ongoing reporting of barcoding procedure compliance (scan rate) and patient ID errors to nursing management. - Duration: 1/1/2009- 6/30/2009 - Training: Education of nursing staff on new practice guidelines includes one-on-one nursing educators; placing “scan only” on each meter, and laminated scanning guidelines cards were attached to each meter tote. - Staff/Other Resources: Not reported - Cost: Not reported	- Outcome Measure: Patient ID error rate: Monthly # of misidentified patients/ total glucose POCTs * Monthly avg. scan rate: # of patient wristbands scanned / Total POCT glucose" Avg scan rate 1-6/2009) = 96.7% Avg scan rate (1/2004) = 31.8% - Recording method: Occurrence log	- Pretest-Posttest - Findings/Effect Size: Patient ID error rate Baseline (1/2004): 2.9% Post (6 mos. 2009): 0.5% ➢ OR = 5.94 (CI,5.26 – 6.71) - Stat. Significance/ Tests: Not reported - Results/ Conclusion Bias: Pre and post comparison practices include POCT barcoding; results show effect of improving implementation of POCT barcoding as reflected in average scan rates

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Sample - Comparator - Study Bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Quality Rating (10 point maximum): 8 (Good) **Effect Size Magnitude Rating : Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2;** **- Study design/time period may introduce bias affecting results**	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 2;** **- Appropriateness of statistical analysis: Does not provide data sufficient to allow/verify calculation of an effect size (sample size not reported)**
- Unpublished POCT - Barcoding A [1]: Anonymous - 2009 - LMBP Network - LMBP Submission Form - [1] Midwest Academic Medical Center, Minnesota, USA - Funding: In-house, quality management project	- Design: Before-after - Facility/ Setting: Midwest- MN Pathology Laboratory, Teaching Hospital; >n 300 beds; >1 million tests/yr. - Time Pd: 1-5/2009; Pre: 1/1/09 - 3/31/09; Post: 4/1/09 – 4/28/09 - Population/Sample: all inpatient and outpatient bedside glucose; Annual glucose POCT performed (2008): 247,000 - Comparator: Manually verify patient armband to glucose work list - Study bias: Short measurement pd of candidate practice (4/1/09 – 4/28/09)	- Description: POCT Barcoding - Duration: 4/1/2009 – 10/1/2009 - Training: Training needs are modest – time to train all glucose users on new process; support for barcode accessories; on-going training on policy and process updates. - Staff/Other Resources: Time to identify & test new armband-materials barcodes, develop barcodes for user-id and maintain process - Cost: Not reported	Outcome Measure: Patient ID error rate – % of POCT glucose results reported on the wrong patient (reported quarterly) - Recording Method: Occurrence log	Pretest-Posttest Findings/Effect Size: Patient ID error rate Pre (Jan – Mar 2009): 0.097%; Post (April 2009): 0.061% 37% reduction ➢ OR = 1.59 (CI, 1.06 – 2.39) *(Note: Denominator of rate estimated from data provided by author—not explicitly stated.)* - Stat. Significance/ Test(s): No report - Results/Conclusions Bias: Authors state Std Dev and means reported – none provided; denominator info not provided to replicate results. No power or stat. test reported
Quality Rating (10 point maximum): 7 (Fair) **Effect Size Magnitude Rating: Moderate** **(Relevance: Less Direct)**	**Study (3 pts maximum): 2;** **- Study time period may not be representative of practice results**	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 1;** **- Measurement period insufficient** **- Appropriateness of statistical analysis: Sample size estimated by reviewers.**

Practices associated with timely and accurate CRITICAL VALUE REPORTING

Laboratory Medicine Best Practices

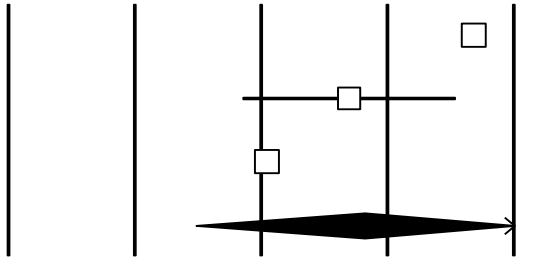
Consensus Ratings 2009

TOPIC AREA: Accurate and Timely Reporting of Critical Values

Practice: Call Centers	Study Quality Rating						Effect Size Rating	Overall Consistency	Overall Strength of Body of Evidence
	Study	Practice	Measures	Results	Total	Rating			
Saxena et al. 2005	2	2	2	1	7	Fair	Substantial	Yes	2 Studies = Fair/Substantial 1 Study = Fair/Moderate 1 Study = Good/Minimal 1 Study = Poor/Minimal (excluded) Suggestive
Unpublished A 2008	2	2	2	1	7	Fair	Moderate		
Providence-Everett 2009	0	1	2	0	3	Poor	Minimal/None		
Geisinger 2009	3	2	1	1	7	Fair	Substantial		
Unpublished B 2009	3	2	2	1	8	Good	Minimal/None		

Critical Value Reporting: Call Centers

Model	Study name	Time point	Statistics for each study			Std diff in means and 95%CI
			Std diff in means	Lower limit	Upper limit	
	Geisinger 2009 *	Blank	1.684	1.635	1.733	
	Unpub Call Ctr A 2 *	Blank	0.697	-0.149	1.543	
	Saxena 2005 *	Blank	0.047	0.000	0.094	
Random			0.814	-0.520	2.148	



-2.00 -1.00 0.00 1.00 2.00

**="Good" Study Quality rating
*="Fair" Quality rating
Boxes proportionate to study size

Favours Comp Favours Call Center

Critical Value Reporting: Call Centers

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Saxena, S (1,2); Kempf, R (1,2); Wilcox, S; Shulman, IA; Wong, L; Cunningham, G; Vega, E, and Hall, S. (2) - 2005 - *Joint Comm J Qual Patient Saf*. [1] Keck School of Medicine, University Southern California, [2] Los Angeles County + University of Southern California Healthcare Network - **Funding:** Self-financed	- **Design:** Cross-sectional - **Facility/Setting: L**A County and Southern Calif Mel Ctr.; Urban, acute care teaching hospital; >700 beds - **Time Pd:** 11/03 – 12/04 Pre: 11/2003 -4/2004 Post: 5/2004 – 12/2004 **- Population/Sample**: All CV lab test notifications for selected inpatient wards and outpatients (not micro); range: 334-700/ mo. Pre: Not reported Post: 4,042 - **Comparator**: Identical standardized call center system w/o IT practice (integrated electronic form) - **Study Bias:** Both practice and comparator include call center practice	- **Description:** Centralized/standardized call center system - Lab tech calls customer service center (CSC) which is responsible for directly communicating CV test result to designated physician by telephone. Utilizes an integrated IT application/database with electronic telephone abstract form filed in the patient record with test results. Physician required to "read-back" patient information for verification. - **Duration:** 14 months (11/03-12/04) - **Training**: 10 hrs. training CSC staff to use system **- Staff/Other Resources:** Interdisciplinary team - **Cost**: 230 hours IT time over 5-month period for development	**Outcome measures:** (1) Time to receipt of CV result in minutes (2) Timeliness of reporting - % CV results reported within 1 hour (3) Timeliness of reporting –% CV results reported within 15 min - **Recording method**: Automated data tracked through IT interface	Pretest-Posttest & Descriptive **- Findings/Effect Size:** (1): Monthly average CV lab test notification time: Pre: 38 minutes Post: 10 minutes (2) Noncomparative: "For May 2004-December 2004, almost all (99%-100%) notifications were completed within one hour"* (3) Noncomparative: "For May 2004-December 2004, 79%-83% of notifications were completed within 15 minutes."* * cited directly from source document; no additional data provided. **- Stat. Significance/Test(s):** Not reported. **- Results/Conclusion Bias:** No data sources provided for outcomes reported; no comparison period sample size reported.
Quality Rating (10 point maximum): 7 (Fair) **Effect Size Magnitude Rating (Relevance: Direct): Substantial**	**Study (3 pts maximum): 2;** - Study design may introduce bias that would affect results.	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 1**; - Appropriateness of statistical analysis: Does not provide data sufficient to allow/verify calculation of an effect size (sample size not reported)

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Unpublished - Call Center: Study A - 2008 - LMBP Network - Eastern USA - **Funding:** Self-financed	- **Design:** Before-after - **Facility/Setting:** Large urban academic medical center in Mid-Atlantic U.S.; > 600 beds; > 32,000 inpatients/yr. and 300,000 outpatients - **Time Period:** 3/08-5/08 Pre: 3/26/2008-4/27/08 (33 days) Post: 4/28/2008- 5/28/2008 (31 days) **- Population/Sample**: No sample size reported. Approximately 200 CV calls/day – likely inpatient only - includes all CV test results within time period - **Comparator**: Computer call queue software tracks time to report CVs to licensed caregivers without using call center. - **Study Bias:** None	- **Description:** Call center Operates 24 hrs./7 days/wk. with a 1 hour target threshold for all CV calls. Lab-certified CV test results go into call center computer queue for its staff to call licensed caregivers. Call Center staff asks caregiver to read-back the results, and documents the read-back in the computer system. Utilizes escalation procedure to identify patient caregiver. - **Duration:** 1 month (Practice initiated on 4/28/08) - **Training**: Not discussed **- Staff/Other Resources:** Staffed by 1-3 medical technologists per shift - **Cost**: Not reported.	**Outcome measures:** (1) Timeliness of reporting - % daily CV results reported within 1 hour (2) Time to receipt of result - Average daily time (min.) per CV test result notification (i.e., to report to licensed caregiver) **- Recording method:** Person making call asks caregiver to read-back results. The read-back is recorded in the computer system, which tracks time from when result certified until caregiver notified (CV "TAT")	Pretest-Posttest: **- Findings/Effect Size:** (1) % CV results reported within 1 hour: Pre: 76.7% daily average (SD: 13.74; Variance: 188.69; Range: 37.5 - 95.3% daily) Post: 92.1% daily average (SD: 5.35; Variance: 28.62; Range: 71.6 – 99% daily). (2) Noncomparative: Pre-Call Center only (3/26–4/21/08): Avg. daily CV notification time: 46.5 minutes (SD: 25.53; Range: 21 – 157); removing the single 157 min. outlier: 42.1 minutes (SD 12.25) - **Stat. Significance/Test(s):** Not reported. **- Results/Conclusion Bias:** No comparisons available on differences between areas where call center was and was not implemented.
Quality Rating (10 point maximum): 7 (Fair) **Effect Size Magnitude Rating (Relevance: Direct): Moderate**	**Study (3 pts maximum): 2;** **-** Study sample may not be representative of practice; - call center not implemented hospital-wide, no information on population	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 1**; - Sample sufficiency: Measurement period may be insufficient to allow robust estimate of impact - Appropriateness of statistical analysis: Does not provide data sufficient to allow/verify calculation of an effect size (sample size)

Evidence Summaries: Critical Value Reporting

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Providence Regional Medical Center, Everett, WA, USA - 2009 - LMBP Network Submission - Funding: Self-financed	- Design: Observational - Facility/Setting: Providence Regional Medical Center in Everett, WA; > 300 beds; >1 million tests/yr. - Time Period:10/08 -6/09 (3 consecutive calendar quarters) 2008- 4th qtr: 10/08-12/08 2009 -1st qtr: 1/09-3/09 2009- 2nd qtr: 3/09-6/09 - Population/Sample: Call center: 108 outpatient only hospital CV calls Comparator: 1,162 inpatient only hospital CV calls - Comparator: Inpatient CV test results communicated by laboratory techs -Study Bias:	- Description: Critical values results communicated by client services call center to designated physician or clinic staff. - Duration: 10/08-6/09 - Training: Client services staff trained on call center management software - Staff/Other Resources: Not reported - Cost: Not discussed	Outcome Measures: Timeliness of reporting - % CV results reported within 15 min. - Recording Method: Occurrence log	Comparison: Call Center (outpatient) v. Techs (inpatient) - Findings/Effect Size: Timeliness within 15 min- 2008-4th qtr: Call Center: 97% (n=29); Techs: 99.8% (n=427) 2009 – 1st qtr Call Center: 97% (n=32); Techs: 98% (n=329) 2009- 2nd qtr: Call Center: 60% (n=47); Techs: 99% (n=406) - Stat. Significance/Test(s): Not discussed - Results/Conclusion Bias: Sample selection may explain unfavorable direction of results
Quality Rating (10 point maximum): 3 (Poor) Effect Size Magnitude Rating (Relevance: Direct): Minimal/None	**Study** (3 pts maximum): 0; - Samples for the practices are sufficiently different to clearly nullify generalizability of the results – small number of outpatient only CV calls for call center vs. large number of inpatient only CV calls for comparator	**Practice** (2 pts maximum): 1; - An important aspect of implementation not well-described; staffing not reported	**Outcome measures** (2 pts maximum): 2	**Results/findings** (3 pts maximum): 0; - Sample sufficiency: Statistical power is not discussed AND the sample is likely too small to allow a robust estimate of the impact of a practice - Appropriateness of statistical analysis: Does not provide data sufficient to allow/verify calculation of an effect size

Evidence Summaries: Critical Value Reporting

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Geisinger Medical Center, Danville, PA, USA - 2009 - LMBP Network Submission - Funding: Self-funded	- Design: Before-After - Facility/Setting: Geisinger Medical Center, Danville, PA; teaching hospital with > 300 beds; >1 million tests/yr. - Time Period: 1/2006--6/09 Pre: 2006 (12 mos.) Post: 1- 6/2009 (6 mos.) - Population/Sample: Avg. 70 CV calls/day to inpatient units and ER. All CVs excluding Anatomic Pathology reported for GMC testing population; Post: 12,306 CV calls; Pre: sample size not reported. - Comparator: Passive system used by bench technologists using a written call log with no readback verification. - Study Bias: None noted	- Description: Call center operates 24 hrs./7 days/wk. and is staffed by21 FTEs. A centralized Client Service Contact Center with an integrated software application make critical value calls directly to a licensed practitioner who can take action on critical values. The Call Center must also verify and document readback of the critical value. The time interval is measured from the identification of the verified critical value to the receipt by the responsible licensed care giver. - Duration: 1/07 - practice ongoing - Training: Education materials provided - Staff/Other Resources: Call Center staff - Cost: Not discussed	Outcome Measures: (1) Timeliness of reporting - % CV results reported within 30 min interval from identification of the verified critical result to acknowledgement by responsible licensed caregiver. - Recording Method: Vendor occurrence/monitoring; No reliable method of tracking comparator rates (2006) as no monitoring system was in place to ensure that the results were given to care providers nor was there documentation of the readback of results.	Pretest-Postest - Findings/Effect Size: (1) % CV results reported within 30 min to responsible licensed caregiver Pre (2006): 50% Post (2009): 95.5% - Stat. Significance/Test(s): Not reported - Results/Conclusion Bias: Data collected during notably different time periods (2006 and 2009); data not provided to support findings or statistical analysis
Quality Rating (10 point maximum): 7 (Fair) **Effect Size Magnitude Rating (Relevance: Direct): Substantial**	**Study (3 pts maximum): 3**	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 1**	**Results/findings (3 pts maximum): 1;** **- Appropriateness of statistical analysis: Compares two practices with estimates based on data from notably different time periods** **- Data insufficient to allow/verify calculation of an effect size (no sample sizes reported)**

Evidence Summaries: Critical Value Reporting

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Unpublished - Call Center: Study B - 2009 - LMBP Network Submission - Western USA - Funding: Self-financed	- Design: Time-series - Facility/Setting: Western USA, Large Health Maintenance Organization (HMO) Laboratory; > 300 beds; >1 million tests/yr. - Time Period: 4-7/2009 April-May & June-July 2009 - Population/Sample: A sample of 500-750 CV test results/mo; study population was CVs for routine outpatient laboratory work. Inpatient or STAT excluded. No sample sizes reported. - Comparator: None: Note: Original CV protocol: Regional Lab notified collection laboratory which then notified provider. -Study Bias: None	- Description: Call center operates 24 hrs./7 days/wk. 24/7 Adult & Advice Call Center that was already staffed with Advice RNs and Call Center MDs who were rotating Emergency Physicians. Two tracks were created, one for INR CVs and one for all other lab CVs - Duration: 3/09 - ongoing - Training: Not discussed - Staff/Other Resources: Staffing level unknown; skilled nursing staff call center; lab assistants occasionally assist in notification. - Cost: Not discussed	Outcome Measure: Timeliness of reporting – % CV results reported within 1 hour - Recording Method: Internal quality control instrument; audit of electronic medical record	Time series - Findings/Effect Size: Timeliness of reporting (within 1-hr) N = 550-750 CVs monthly (2009) Time 1 (April-May 2009): 647 CVs notified within 1 hour/ 650 CVs monthly = 99.5% (*estimate - based upon range: 1-6 not reported) Time 2 (June-July 2009): 650 CVs notified within 1 hour/ 650 CVs monthly = 100% - Stat. Significance/Test(s): Not reported - Results/Conclusion Bias: "Workflow change does not account for rare human factor incident where lab personnel may forget to notify CV immediately (within 1 hr) and the CV is caught at the end-of-shift"
Quality Rating (10 point maximum): 8 (Good) **Effect Size Magnitude Rating (Relevance: Direct): Minimal/None**	**Study (3 pts maximum): 3**	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 1;** **- No statistical significance/tests performed** **- Data do not permit effect size calculation**

Laboratory Medicine Best Practices

Consensus Ratings 2009

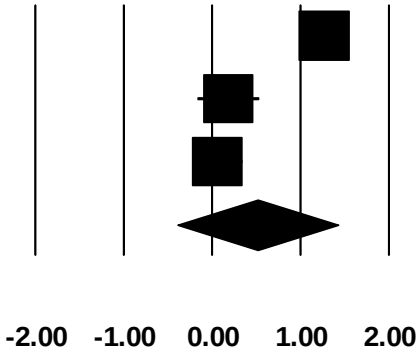
Critical Value Reporting: Automated Notification

Practice: Automated Notification	Study Quality Rating						Effect Size Rating	Overall Consistency	Overall Strength of Body of Evidence
	Study	Practice	Measures	Results	Total	Rating			
Kuperman et al. 1999	2	2	2	1	7	Fair	Minimal/None	Yes	1 Study = Good/Substantial 2 Studies = Fair/Minimal 1 Study = Good/n.a. Suggestive
Park et al. 2008	0	2	2	2	6	Fair	Minimal/None		
Tate et al. 1995	3	2	2	1	8	Good	Not Available		
Piva et al. 2009	3	2	2	1	8	Good	Substantial		

Critical Value Reporting: Automated Notification

Std Effect Est -- Automated Notification to Reduce Lab CV Notification Time

Study name	Outcome	Statistics for each study		
		Std diff in means	Lower limit	Upper limit
Piva 2009 **	% reported < =1 hr	1.269	1.167	1.371
Park 2008 *	Clinical Resp Rate	0.184	-0.168	0.535
Kuperman 1999 *	Time to Resolution	0.058	-0.228	0.344
		0.513	-0.397	1.423



**="Good" Study Quality rating
*="Fair" Quality rating
Boxes proportionate to study size

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) -Results/Conclusion Bias
Kuperman GJ [1,2]; Teich JM [1,2]; Tanasijevic MJ [2]; Ma'Luf N [2]; Rittenberg E [2]; Ashish Jha MA [2]; Fiskio J [1]; Winkelman J [2]; Bate, DW [1,2] - 1999 - *Journal of the American Medical Informatics Association* [1] Partners HealthCare Systems, [2] Harvard Medical School **- Funding:** Partly from research grant (R01 - Agency of Health Care Policy and Research)	- **Design:** Randomized Controlled Study - **Facility /Setting:** Brigham & Woman's Hospital; 720 bed tertiary-care hospital in Boston, MA, - **Time Period:** 12/1994-10/1995 (Medical: 12/1994–1/1995: 2 months; Surgical: 9/1995 – 10/1995: 2 months) - **Population/Sample:** 178 subjects; 192 tests; 4 laboratory tests with critical values and/or alert situations - **Comparator**: Same as practice without alert to physician pager (i.e. alert goes to patient floor computer screens) and with CV results telephoned by lab technologists to patient floor (nursing staff). - **Study Bias:** None	- **Description:** Automated notification - system generates an alert to a digital pager of the patient's covering physician results - **Duration:** 6 months - **Training**: Not discussed **- Staff/Other Resources:** Computer technicians, physician, nurses, unit secretary, telephone operator, reviewers and lab supervisor/manager; clinical alerting system, digital pager, computer workstation - **Cost**: Not reported	**Outcome Measures:** (1) Time to Treat (TTT) - Time interval from the filing of the alerting result to the ordering of appropriate treatment (2) Time to Resolution (TTR) - Time interval from the filing of alerting result to the arrival time in the laboratory of a bedside test demonstrating the alerting condition was no longer present - **Recording method**: Occurrence log, Chart review	- Comparison (RCT) - **Findings/Effect Size:** (1) TTT: Total N = 97; Practice Mean time: 3.4 hours (SD = 8.0; Median =.07; n = 43). Comparator Mean time: 3.3 hours (SD = 7.4; median = 1.1; n = 54) (2) TTR: Total N = 94 Practice Mean time: 12.8 hours (SD = 15.4; median = 7.0; n = 40). Comparator Mean time: 13.7 hours (SD = 14.5; median = 8.1; n = 54) - **Stat. Significance/Test(s):** (1) TTT: Student T-test P-value = 0.59; d = -0.013; statistical power < 80% (2) TTR: Student T-test P-value = 0.68; d = 0.060; statistical power < 80% **- Results/Conclusion Bias:** Favorable conclusions not supported; but instead are contradicted by reported findings. Authors note that differences are not significant but focus on direction of effect.
Quality Rating (10 pt max): 7 (Fair); Effect Size Magnitude Rating: Minimal/None (Relevance: Direct)	**Study (3 pts maximum): 2** **-** Results may not be generalizable: Tertiary-care population	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 1** - Results not clearly attributable to the practice (p-values); conclusions not supported by work

Evidence Summaries: Critical Value Reporting

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) -Results/Conclusion Bias
Park H [1]*, Min WK [1], Lee W [1], Park H [2], Park CJ [1], Chi HS [1], Chun S [1] - 2008 - *Annals of Clinical & Laboratory Science.* [1] Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; [2] Kangbuk Samsung Hospital,Sungkyunkwan University School of Medicine, Seoul, Korea; * Current affiliation: Catholic University of Korea, College of Medicine, Seoul, Korea **- Funding**: Korea Health Industry Dev Institute.	- **Design:** Before-after - **Facility/Setting:** 2,200 bed tertiary care urban, academic medical center, Seoul Korea - **Time Period:** Two 12-month study periods 1/01 – 12/05: Pre:1/1/01 -12/31/01 Post: 1/1/05 - 12/31/05 - **Population/Sample:** Serum potassium alert values – Pre: 121 alert calls for 2001: ICU: 56; general wards 65 Post: 96 alert calls for 2005 ICU: 31; general wards 65 - **Comparator**: Lab tech phones nurses on inpatient floor to notify of patient CV. Nurse then informs physician of patient of CV result. Call documented in lab & rcvng loc - **Study Bias:** None	An automated alerting system which involves the use of a computerized database (i.e., HIS; LIS) -Text messages w/ patient information and test result is transmitted to appropriate physician via PDA phones. - **Duration:** 1/1/05 -12/31/05 - **Training:** Not discussed **- Staff/Other Resources:** Lab technicians, nurses, physicians; Computer software and PDA phones for all physicians - **Cost:** not reported	**Outcome Measures:** (1) Time to receipt - Time interval in minutes from dispatching critical value result alert to acknowledgement by responsible caregiver (2) Clinical response rate (CRR) - response rate defined as the frequency of clinical responses divided by total # critical value alerts - **Recording method:** Occurrence log – (1) LIS – (2)	Pretest-Posttest **- Findings/Effect Size:** (1)Time to receipt Overall mean decreased by 40.8% Pre (2001): 343.3 min (SD = 369.6) Post (2005): 203.2 min. (SD = 294.1). "Median mean" (2001): 213.0 min; "Median mean" (2005): 74.5 min P<.001; d=-0.419 (2)CRR - Overall increase: Pre: 73.3% Post: 79.3% General wards - No change (82.3%) ICU - increase: 65.1% (2001) to 73.8% (2005) - **Stat. Significance/ Test(s):** (1) p= 0.190 (overall Mean); p <0.001 ("median Mean"); (2) Overall rate: p=0.265, d = 0.184, OR = 1.40 **- Results/Conclusion Bias:** None
Quality Rating (10 point maximum): 6 (Fair) **Effect Size Magnitude Rating: Moderate (Relevance: Direct)**	**Study (3 pts maximum): 0** **-** The time period and sampling methods likely to introduce a study bias substantially affecting results; 4–yr gap with relatively small sample - Results are unlikely to be generalizable to other settings given large proportion ICU	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 2** - Appropriateness of statistical analysis: Compares two practices and their estimates are based on data collected during notably different time periods (2001 v. 2005)

Evidence Summaries: Critical Value Reporting

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) -Results/Conclusion Bias
Tate KE [1], Gardner RM [1], and Scherting K [1] - 1995 - *Proceedings from the* ***ANNUAL SYMPOSIUM*** *on* ***COMPUTER*** *Applications in* ***MEDICAL CARE*** [1] LDS Hospital, Salt Lake City, Utah **- Funding:** Not reported	- **Design:** Observational - **Facility/Setting:** Private 520-bed tertiary-care facility part of Intermountain Health Care Hospital System; LDS Hospital, Salt Lake City, Utah; 48-bed adult unit nursing division - **Time period:** 13-weeks (10/1/93-1/21/94) - **Population/Sample:** CV alerts for 335 inpatients (representing 497 critical values or tests) - **Comparator**: None (Previous practice: laboratory staff call nursing unit by telephone when critical test results appear.) - **Study Bias:** None	- **Description:** Automated notification system - **Duration:**10/1/1993 – 1/21/1994 - **Training**: Not discussed **- Staff/Other Resources:** Laboratory staff and nurses - **Cost**: not discussed	**Outcome Measure:** (1) Time to receipt -Time interval (in min.) from dispatching CV result alert to acknowledgement by responsible caregiver (76% of all alerts acknowledged by primary care nurses) (2) Timeliness of reporting - % CV results reported within various time intervals - **Recording method:** HIS – computerized occurrence log	- Non-comparative - **Findings/Effect Size:** (1) Time to receipt: Average weekly alert acknowledgment time: 38.6 min. (2) Timeliness of CV reporting: 51% within 12 min. (4.2 min avg.), 81% within 1 hr. 95% within 2 hrs. - **Stat. Significance/ Test(s):** None reported **- Results/Conclusion Bias:** None
Quality Rating (10 point maximum): 8 (Good) **Effect Size Magnitude Rating: Not available** **(Relevance: Direct)**	**Study (3 pts maximum): 3**	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 1** - Appropriateness of statistical analysis: Does not provide data sufficient to allow/verify calculation of an effect size - Statistical power not discussed; sample may be too small for a robust estimate

Evidence Summaries: Critical Value Reporting

Bibliographic Information - Author (s) - Yr Published - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) -Results/Conclusion Bias
Piva, E [1], Sciacovelli, L [1], Zaninotto, M [1], Laposata, M [2], Plebani, M [1]- 2009 - *American Journal of Clinical Pathology* [1] Department of Laboratory Medicine, Padua University School of Medicine, Padua, Italy; [2] Vanderbilt University Hospital, Nashville TN **- Funding:** not reported	- **Design**: Before-after - **Facility/Setting**: University of Padua, Academic medical center, > 300 bed inpatient hospital; Annual test volume: >1,million - **Time period:** 1/1/07 – 2/28/08; Pre: 1/1/07-12/31/07 (1 year) Post: 1/1/08-2/28/08 (2 mos.) - **Population/Sample:** All critical values Pre: 7,320 CVs (4,394 routine inpatient; 2,926 emergency; 1,323 outpatient) Post: Not reported - **Comparator:** Telephone-only CV notification system - **Study Bias:** None	- An automated alerting system which involves the use of a computerized database (i.e., HIS; LIS) of test results. The alert message flashes on the monitor until the physician or a nurse in charge of notification confirms that the message has been received; the flashing alert is stopped after 60 minutes. - **Duration:** 1/1/08 – 2/28/08 - **Training:** Not discussed **- Staff/Other Resources:** physicians, clinical pathologist, nurses, laboratory - **Cost:** Not discussed	**Outcome Measures:** (1) Timeliness of reporting - % CV results reported within 1 hr:; # unsuccessful notifications w/in 1 hr / total # of CVs (2) Time to receipt: - Time from detection of CV in minutes to acknowledgement by responsible clinician - **Recording method**: register (pre-implementation) and HCIS (health care information system) (post-implementation)	- Pretest-Postest **- Findings/Effect Size:** (1) % Reported within 1 hour Pre: >50% "unsuccessful" Post: 10.9% (2) Time to receipt Pre: Average 30 min; Post: Average 11 min - **Stat. Significance/ Test(s):** None reported **- Results/Conclusion Bias:** No post sample data (numerator or denominator); sample size only for pre-practice period.
Quality Rating (10 point maximum): 8 (Good) **Effect Size Magnitude Rating: Substantial (Relevance: Direct)**	**Study (3 pts maximum): 3**	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 1** - Appropriateness of statistical analysis: Does not provide data sufficient to allow/verify calculation of an effect size - Sample sufficiency: Number of post period subjects not reported.

Practices associated with reduction of BLOOD CULTURE CONTAMINATION

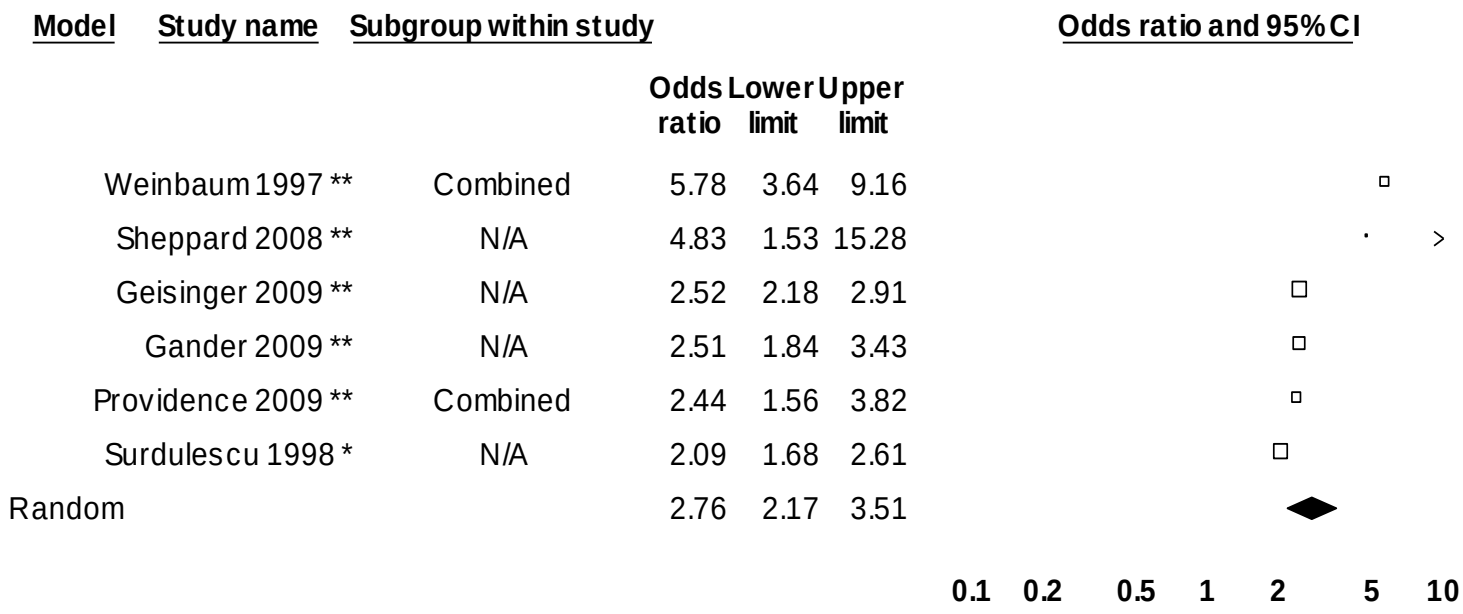
Laboratory Medicine Best Practices

Consensus Ratings 2009

TOPIC AREA: Blood Culture Contamination Consensus Ratings

Practice: Dedicated Phlebotomy Teams	Study Quality Rating						Effect Size Rating	Overall Consistency	Overall Strength of Body of Evidence
	Study	Practice	Measures	Results	Total	Rating			
Gander et al. 2009	2	2	2	3	9	Good	Substantial	Yes	5 studies = good/substantial 1 study = fair / substantial High
Sheppard et al. 2008	1	2	2	3	8	Good	Substantial		
Surdulescu et al. 1998	1	1	1	2	5	Fair	Substantial		
Weinbaum et al. 1997	2	2	2	3	9	Good	Substantial		
Providence-Everett 2009	2	2	2	2	8	Good	Substantial		
Geisinger Wyoming 2009	2	2	2	3	9	Good	Substantial		

DEDICATED PHLEBOTOMY TEAMS



**="Good" Study Quality rating
*="Fair" Quality rating
Boxes proportionate to study size

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Gander RM (1,2), Byrd L (2), DeCrescenzo M (3), Hirany S (2), Bowen M (2), Baughman J (2) - 2009 - Journal of Clinical Microbiology - [1] Dept of Pathology, University of Texas Southwestern Medical Center, Dallas Texas [2] Dept of Pathology, Parkland Health and Hospital System [3] Dept of Performance Improvement, Parkland Health and Hospital System - Funding not reported	- **Design:** Cohort Study (Groups defined by predictor) - **Facility/Setting:** Parkland Memorial Hospital - 968 bed tertiary care teaching hospital, Dallas, Texas - **Time Period:** 12/1/2006-12/31/2007 (data collected for 5 separate months over a 13- month period) - **Population/Sample:** 3,662 blood cultures from 2,642 adult patients in ED west: 2,012 blood cultures by phlebotomists; 1650 blood cultures by nonphlebotomy staff **-*Comparator:** Venipuncture by Nonphlebotomy staff - **Study bias:** - ED setting/samples only - Students and other limited experience/skill staff used for comparator	**Description:** Dedicated phlebotomy team assigned to manage all blood collection and specimen activities - **Duration:** 13 months (12/1/2006- 12/31/2007; ongoing afterward) **- Training**: not discussed, but phlebotomist must be certified. **-Staff/Other Resources:** Cultures collected by phlebotomy teams Aerobic and anaerobic bottles with media. **- Cost:** Not reported.	Blood Culture Contamination Rate (BCCR) **-Recording method:** Blood culture data reviewed for 5 separate months (at 3 month intervals) over a 13 month period.	- Comparison (cross-sectional) - **Findings/Effect Size:** **BCCR** Dedicated phlebotomy practice: Overall: 3.1% (62/2012); monthly range: 2.4 to 3.6 Nonphlebotomy: Overall: 7.4% (122/1650); monthly range 6.2 to 10.2% **Effect Size:** OR = 2.51 (CI,1.84 – 3.43) **- Statistical significance/Test(s):** chi-square =34.41 df=1, p<.0001 **- Results/conclusion biases**: None noted
Gander 2009 **Quality Rating (10 point maximum): 9 (Good)** **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2** - Only ED setting/patients/tests; not generalizable due to higher ED BCCRs	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts. Maximum): 2**	**Results/findings (3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Sheppard C (1), Franks N (2), Nolte F(1), Fantz C (1). - 2008 - Am J Clinical Pathology [1] Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA. [2] Department of Emergency Medicine, Emory University, Atlanta, GA. - Funding: Self-funded	- **Design**: Cohort Study **Facility/Setting**: Emory Crawford Long Hospital; Academic Medical Center Atlanta, GA, USA **Time Period**: No dates reported – 9 months (6 months comparator; 3 months practice) - **Population/Sample:** Total 2,854 blood cultures collected in the ED; nonphlebotomist comparator 6-month sample = 2,576; phlebotomy practice 3-month sample 278 Note: 1/4 - 1/3 of all blood cultures in hospital from the ED **- Comparator:** Nonphlebotomy staff working 2:00 pm to 10:00 pm collected 6 months before the intervention. - **Study bias:** Potential patient selection bias – higher acuity for phlebotomist. ED only	**Description**: Phlebotomist dedicated to the ED randomly collected specimens on the weekday evening shift (2:00 pm-10:00 pm) **Duration**: 3 months. – no dates reported **- Training:** Not reported **- Staff:** 1 dedicated lab phlebotomist **-Cost/Other Resources:** Proposed implementation-annually for dedicated phlebotomist in ED (annual blood culture volume of approximately 6,000): 8.4 FTEs at $13.49/h (2.0 FTEs each for day, evening, and night shifts and 2.4 FTEs for weekends and 25% benefits)	Blood Culture Contamination Rate (BCCR) **- Recording method:** BCC data collected quarterly and reported by department and collection personnel identifiers.	- Pretest-Posttest **Findings/Effect Size:** **BCCR** Phlebotomist: 1.1% (3/278 cultures) Nonphlebotomist:: 5.0% (129/2576 cultures) Note: 1.1% for phlebotomist collection was not significantly different from the average phlebotomy rate for the hospital of 1.3%. **Effect Size: OR = 4.83 (CI, 1.53 – 5.28)** **Statistical Significance/Tests** P =.001 -**Results/conclusion biases**: none
Sheppard 2008 **Quality Rating (10 point maximum): 8 Good** **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 1** - Study/practice dates not provided - ED setting only - Practice sample "random" but small relative to total volume	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts. maximum): 2**	**Results/findings (3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
-Ramsook C (1,2), Childers K (2), Cron SG (3), Nirken M (2) - 2000 - Infect Control Hosp Epidemiology [1] Eric Williams Medical Sciences Complex [2] Emergency Medicine, Texas Children's Hospital, Baylor College of Medicine, Houston, TX [3] Academic General Pediatrics, Baylor College of Medicine **-** Funding not reported	**- Design:** Observational Study **- Facility/setting:** Texas Children's Hospital; Houston University-affiliated pediatric emergency room Houston, Texas, USA - **Time period:** 2/1/1999-7/31/1999 **- Population/Sample:** Single blood cultures drawn by phlebotomists from 1,073 patients **- Comparator:** Venipuncture by nurses using three sequential Betadine swabs followed by three sequential alcohol swabs for skin antisepsis **-Study biases:** -Pediatric ED setting - Limited experience/skill staff used for comparator	**Description:** Laboratory phlebotomists using one Betadine swab followed by one alcohol swab for skin preparation **- Duration:** 6 months **- Training:** not discussed, but phlebotomist must be certified. **- Staff/Other resources:** Cultures processed by laboratory phlebotomist, swabs, and alcohol swabs **- Cost:** not reported	Blood Culture Contamination Rate (BCCR) **-Recording method**: Blood culture data reviewed for 6 month period.	- Non-randomized Comparison **Findings/Effect size:** **BCCR** Laboratory phlebotomist: Overall: 2.0% (17/646); 2.6% Nonphlebotomy: Overall: (5/427); 1.2% **- Statistical significance/Test(s):** - Phlebotomist versus not-phlebotomist relation not tested - Statistical power is not discussed **Effect Size: OR = 0.44 (CI,0.16 – 1.20)** **- Results/conclusion biases:** Potential confound if phlebotomists are used to collect samples from younger patients (documented relationship of patient age with BCCR); potential confound from different skin preparation
Ramsook 2000 ***Quality Rating (10 point maximum): 5 Fair** **Effect Size Magnitude Rating: Adverse** **(Relevance: Direct)**	**Study (3 pts maximum): 2** - Only Pediatric ED setting; not generalizable due to high ED BCCRs	**Practice (2 pts maximum): 2**	**Outcome measure (2 pts. Maximum): 1** Clinical histories not obtained, patients on antibiotics not excluded	***Results/findings (3 pts maximum): 0** **-** More rigorous skin prep technique applied by nurses is a potential confounder for lower BCCR rates

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Surdulescu S(1), Utansingh D(1), Shekar R(1) - 1998 - Clinical Performance and Quality Healthcare - Dept. of Internal Medicine, The Division of Infectious Disease, St. Luke's Medical Center, Case Western Reserve University, Cleveland, Ohio - Funding not reported	- **Design:** Retrospective cohort study/Natural experiment with nested case-control study (Groups defined by outcome). **Facility /Setting:** St. Luke's Medical Center, Case Western Reserve University, Cleveland, Ohio - **Time period:** 1/1/1993-12/1/1995 - **Population/Sample:** Observational study: Sample size not explicitly stated. 6900 phlebotomy venipunctures in 1995. Case-control study 23 patients contaminated blood cultures matched with 23 patients with negative blood cultures[1] **Comparator:** Non-phlebotomy staff blood draws with prep kit - **Study bias:** none noted	Dedicated phlebotomy team assigned to manage all blood collection and specimen activities. Blood draws with prep kit. **Duration:** 24 months - 1/1/1993-12/1/1995 **- Training:** not discussed, but phlebotomist must be certified. - **Staff:** not discussed - **Cost:** not reported	Blood Culture Contamination Rate (BCCR) - **Recording method:** Not described for cohort study. Case-Control: Chart reviews by physicians	- Comparison (Case control) **Findings/Effect size:** **BCCR** Dedicated Phlebotomy practice: 2.6 %From Jan 1993-Oct 1993 Nonphlebotomy: 5.6%From Jan 1993-Oct 1993 Phlebotomy teams were eliminated From Nov 1993-Dec 1995 Overall: 4.5 % to 5.8% in 1994 and 5.3% in 1995; p=.001 **Effect Size: OR = 2.09 (CI,1.68 – 2.61)** [2] - Statistical test: Chi-square, 2 df. - Statistical power is not discussed. **- Results/conclusion biases:** none noted
***Quality Rating (10 point maximum: 5 Fair** **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 1** -Cohort sample size not explicitly stated - Use of commercial prep kit potentially limits generalizability	**Practice (2 pts maximum): 1** -Staffing qualifications for practice and comparator not well-described.	**Outcome measure: (2 pts maximum): 1** - As QC, it is unknown whether same physicians drawing the cultures were reviewing the charts.	**Results/conclusion Biases: (3 pts maximum) 2** - Denominators for proportions not reported

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Weinbaum FI (1, 2), Lavie S (3), Danek M (2), Sixsmith D (4), Heinrich GF (5), Mills SS (5). - 1997 - Journal of Clinical Microbiology - New York Hospital Medical Center of Queens, Flushing, New York [1] Dept of Surgery [2] Dept of Quality Management [3] Dept of Pathology [4] Dept Emergency Medicine [5] Administration - Funding not reported	**- Design:** Nonrandomized prospective intervention trial **- Facility/Setting:** New York Medical Center Hospital of Queens; 487 bed Community Hospital Center Flushing, NY, USA **- Time period:** No dates reported- 9 months (6 months comparator; 3 months practice) - **Population/Sample:** 956 Blood cultures drawn by Blood Culture Collection Team (BCT) with prep kits - **2 Comparators:** (1) House staff conducting draws without prep kits. (2) House staff conducting draws with prep kits. - **Study bias:** none noted	**- Description:** Blood Culture Team (BCT) made up of three full-time phlebotomists with prep kits. **- Duration:** 9 months- no dates reported - **Training:** not discussed, but phlebotomist must be certified - **Staff/Other resources:** Three full-time phlebotomists using prep kits; aerobic and anaerobic bottles - **Cost**: not reported	Blood Culture Contamination Rate (BCCR) - **Recording method**: Internal quality control instrument	- Comparison between independent groups -**Findings/Effect size:** **BCCR** House staff with prep kit 4.8%, (.016) House staff without prep kit 8.4% (.014) DPT with prep kit 1.1% (.004) **Effect Size: OR = 4.34 (CI,1.82 – 10.36)** **- Statistical significance/Tests:** Mantel Haenszel Chi-square = df=1, p<0.001; For house staff with prep kit vs. house staff without prep kit, P=.173. **- Results/conclusion biases:** none Noted
Quality Rating (10 point maximum): 9 Good **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2** -No dates for time period specified, however duration of practice described; concomitant use of second practice (commercial prep kit) potentially limits generalizability.	**Practice (2 pt maximum): 2**	**Outcome measures(2 pts maximum): 2**	**Results/findings (3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Unpublished - Providence Regional Medical Center – Everett -2009 -LMBP Network - Funding: In house - as part of ongoing Patient Safety/Quality Indicators from 2005 to present.	**- Design:** Before-after **- Facility/Setting:** Providence Regional Medical Center – Everett; Non-teaching hospital; >300 beds Everett, Washington, USA **- Time period:** 1/1/2005-7/31/2009 - **Population/Sample:** Currently process ~1900 blood cultures per month. In 2005 the volume was ~1100. - ***Comparator**: Venipuncture and line draws by Non-phlebotomy staff **- Study bias:** - ED and CCU setting/Samples Based on data points for 2 months – cannot assess long-term variation	**- Description:** Blood cultures collected by Lab phlebotomists; training coupled with individualized feedback to personnel performing blood draws. **- Duration:** 1/1/2005; ongoing afterward **- Training:** Data sharing with Nursing and medical staff; education on collection technique. -**Staff/Other Resources:** ~400 hours of phlebotomy support to draw 700+ BC/mos collected by nursing staff. - **Cost:** not reported	Blood Culture Contamination Rate (BCCR) - **Recording method:** Internal quality control instrument linked to laboratory information system	- Comparison **Findings/Effect size:** **BCCR** <u>Jan – Mar, 2005</u> -Lab Phlebotomy Team 3.0% -Non Lab personnel 6.0% May – Jul, 2009 -Lab Phlebotomy Team 0.9% - Non Lab personnel 3.1% **Effect Size: OR = 2.44 (CI, 1.56 – 3.82)[1]** - Statistical power is not discussed **- Results/conclusion bias:** - Possible confound by change in prevalence of non-phlebotomist use of venous catheters for blood specimens over 4.5 year study period
Quality Rating (10 point maximum): <u>8 Good</u> **Effect Size Magnitude Rating: <u>Substantial</u>** **(Relevance: Direct)**	**Study (3 pts maximum): <u>2</u>** - Generalizability limited because of concomitant use of second practice (individualized monitoring & feedback).	**Practice (2 pts maximum): <u>2</u>**	**Outcome measures (2 pts maximum): <u>2</u>**	**Results/findings (3 pts maximum): <u>2</u>** - Change in prevalence of line draws by non-phlebotomists is a confounder for results

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Unpublished - Geisinger Wyoming Valley Hospital - 2009 - LMBP Network - Self-funded	**- Design:** Observational – time series **- Facility/Setting:** Geisinger Wyoming Valley Hospital; Inpatients, ER, Urgent Care and Outpatients, Wilkes-Barre PA **- Time period:** 1/2006-9/2009 (45 months) **-Population/Sample:** All inpatients, Emergency Department, Urgent Care and Outpatients. Approximately 98% of the blood cultures collected are inpatient and Emergency Department On average, 780 blood cultures collected at site monthly in 2009. 73% by phlebotomists; total estimated sample size approximately 35,100 (45 x 780); 9,360/year **- *Comparator:** Venipuncture and line collections (A-line, Pic line, dialysis etc…) by nonphlebotomy staff **- Study bias:** Inclusion of line draws in comparator may bias estimated difference in rates.	**- Description:** Blood cultures collected by laboratory phlebotomists. All lab draws are peripheral collections. **- Duration:** 45 months 1/2006-9/2009; ongoing **- Training:** not discussed, but phlebotomist must be certified. **-Staff/Other Resources:** Cultures collected by laboratory phlebotomist. **- Cost:** not reported	**Description:** Blood Culture Contamination Rate (BCCR) Laboratory phlebotomist Nonphlebotomy [2] **- Recording method:** All blood cultures designated by Microbiology to be contaminants.	Observational – time series **Findings/Effect Size:** BCCR **Laboratory phlebotomist**: Annual monthly averages 2006: 2.6% 2007: 1.8% 2008: 1.3% 2009 (9 months): 1.5% Overall monthly average: (45 months): 1.8% **Non-phlebotomy:** Annual monthly averages 2006: 4.9% 2007: 4.3% 2008: 3.9% 2009 (9 months): 3.9% Overall monthly average (45 months): 4.3% **Effect Size: OR = 2.52 (CI, 2.18 – 2.91)[4]** **- Statistical significance/Tests(s):** Monthly averages reported; statistical analysis/testing not reported, but analysis by reviewers obviates need. **Results/conclusion biases:** None
Geisinger 2009 **Quality Rating (10 point maximum): 9 (Good)** **Effect Size Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2** - lack of monthly sample size data	**Practice(2 pts maximum): 2**	**Outcome measure (2 pts maximum): 2**	**Results/findings (3 pts maximum): 3**

Laboratory Medicine Best Practices

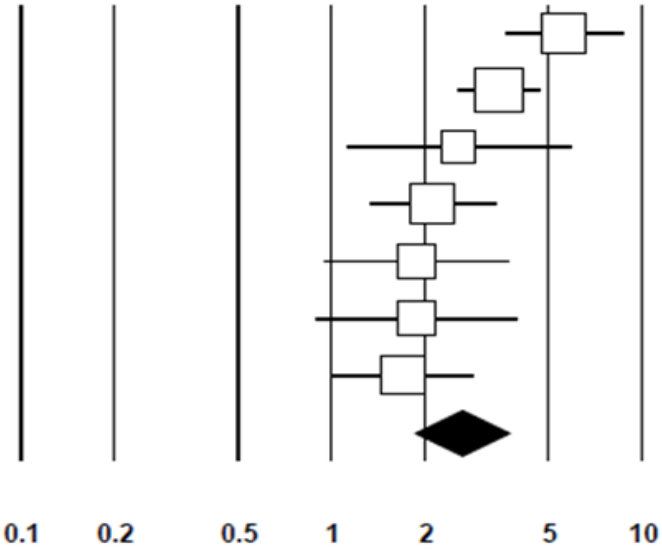
Consensus Ratings 2009

VENIPUNCTURE VS. INTRAVENOUS CATHETER STUDIES

Practice: Venipuncture (vs. Catheter) Collection Site	Study Quality Rating						Effect Size Rating	Overall Consistency	Overall Strength of Body of Evidence
	Study	Practice	Measures	Results	Total	Rating			
Desjardin et al. 1999	3	1	2	3	9	Good	Moderate	Yes	3 Studies = Good/Substantial 1 Study = Good/Moderate 2 Studies = Fair / Substantial High
Everts et al. 2001	3	1	1	3	8	Good	Substantial		
Martinez et al. 2002	3	2	1	1	7	Fair	Substantial		
McBryde et al. 2005	3	2	2	3	10	Good	Substantial		
Ramsook et al. 2000	3	2	1	1	7	Fair	Moderate		
Norberg et al. 2003	3	2	1	3	9	Good	Substantial		

VENIPUNCTURE VS. INTRAVENOUS CATHETER STUDIES

Study name	Odds ratio	Lower limit	Upper limit
McBryde 2005**	5.60	3.61	8.69
Norberg 2003**	3.46	2.55	4.69
Martinez 2002*	2.57	1.12	5.89
Everts 2001**	2.12	1.32	3.40
DesJardin 1999**	1.88	0.95	3.74
Beutz 2003**	1.88	0.88	3.99
Ramsook 2000*	1.70	1.01	2.85
	2.63	1.85	3.72



**="Good" Study Quality rating
*="Fair" Quality rating
Boxes proportionate to study size

Boxes proportional to study size.

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Beutz M (1), Sherman G (2), Mayfield J (3), Fraser VJ (4), Kollef MH (1). **- Year:** 2003 **- Publication:** -Chest **- Author Affiliation:** (1) Pulmonary and Critical Care Division, Washington University School of Medicine; (2) Departments of Nursing, Barnes-Jewish Hospital; (3) Infection Control, Barnes-Jewish Hospital (4) Division of Infectious Diseases, Washington University School of Medicine **- Funding:** not reported	- **Design:** Prospective cohort - **Facility/Setting**: Barnes-Jewish hospital; Teaching hospital; ICU > 300 beds St Louis, MO, USA. ≅1,600 ICU admissions annually - **Time Period**: 2/1/2001-10/30/2001 **- Sample:** All 119 patients admitted to ICU with temperatures > 38.3°C and BC specimens drawn from a central vein catheter (CVC) and peripheral venipuncture (PV) within 4 hours of each other. Multiple pairs allowed if drawn > 24 hours apart. 300 BC pairs **- Comparator:** BC drawn by PV, 70% isopropyl alcohol followed by 2% iodine disinfectant **- Biases:** None	- **Description**: Blood cultures drawn from Central Venous Catheters[1] - **Duration:** 8 months **- Training:** not discussed **- Staff:** Blood cultures obtained by critical care nursing staff. **- Cost:** not reported	**- Description:** 3 Outcome measures: (1) BCCR (2) PPV* (3) Specificity* **- Recording method:** Two study physicians blinded to the site of BC classified paired cultures.[2] If both BC positive, accepted as positive; if both BC negative, accepted as negative. Disparate results resolved for true bacteremia.	**- Type of Findings:** Paired comparisons **- Findings:** (1) BCCR: CVC= 20/300 (6.7%); PV = 11/300 (3.7%). Remainder (269/300) correctly identified. (2) PPV: CVC = 58.3% (CI, 44.4% to 72.2%); PV = 66.7% (CI, 50.6% to 82.8%). (3) Specificity: CVC = 92.5% (CI, 89.4% to 95.6%); PV = 95.9% (CI, 93.5% to 98.3%) **Effect Size: OR = 1.88 (CI, 0.88 – 3.99)** **- Statistical Significance/Tests:** - Difference in main effects not tested, CI for results 2 & 3 = no significant differences Catheter vs. PV - Stepwise mult. logistic regression - Statistical power is not discussed **- Results/conclusion bias:** none
Quality Rating **(10 point maximum): 9** **Good** **Effect Size Magnitude Rating: Moderate** **(Relevance: Direct)**	**Study** **(3 pts maximum): 3**	**Practice** **(2 pts maximum): 2**	**Outcome measures** **(2 pts maximum): 2**	**Results/findings** **(3 pts maximum):2** **-** Difference in main effects not tested

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- DesJardin JA (1), Falagas ME (2), Ruthazer R (3), Griffith J (3), Wawrose D (4), Schenkein D (3), Miller K (3), Syndman DR (3) - Year: 1999 - Publication: Annals of Internal Medicine - Author Affiliation: (1) Western Infectious Disease Consultants (2) Athens, Greece (3) New England Medical Center (4) Nashville, TN - Funding: In part by National Research Service Award T32-A107329 training grant from NIH,	- Design: Retrospective cohort - Facility/Setting: New England Medical Center; Academic Medical Center; 100-300 beds; Tertiary oncology ward, Boston, MA, USA - Time Period: 8/1/1994-6/1/1996 - Sample: 551 paired BC cultures obtained from 185 patients with 306 admissions. Paired culture defined as at least one BC clearly labeled as drawn from a central venous catheter and at least one blood sample drawn through a peripheral venipuncture (PV). Blood samples had to be drawn within 4 hours of each other. - Comparator: PV - Study Bias: None noted	- Description: Blood cultures drawn from Central Venous Catheters (CVC) paired blood cultures. - Duration: 22 months - Training: Not discussed - Staff/Resources: Not described - Cost: Not reported	- Description: 3Outcome measures: (1) Blood Culture Contamination Rate (BCCR) – percentage of all blood cultures growing contaminants (2) PPV* (3) Specificity* -Recording method: Retrospective screening of all BCs. Blinded assessments of culture results done by infectious disease experts were used as the gold standard.1	- Type of Findings: Paired Comparisons - Findings: (1) BCCR:CVC = 24/552 (4.3%); PV = 13/552 (2.35%) (2) PPV: CVC= 63% (CI= 50% - 75%); PV = 73% (CI=60% - 86%) (3) Specificity CVC 95% (CI =93% - 97%) PV = 97% (CI=96%-99%) Effect Size: OR = 1.88 (CI, 0.95 – 3.74) - Statistical Significance/Tests: - Bootstrap methods used to adjust for potential clustering around patient, hospital admission, and blood culture.2 - Statistical power is not discussed/NA -Results/conclusion bias: None noted
Quality Rating (10 point maximum): 9 (Good) **Effect Size Magnitude Rating: Moderate (Relevance: Direct)**	**Study (3 pts maximum): 3**	**Practice (2 pts maximum): 1** **- Staffing qualifications for practice not described.**	**Outcome measures (2 pts maximum): 2**	**Results/findings (3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- McBryde, ES(1,2), Tilse, M(2), McCormack, J(2). - Year: 2005 - Publication: Journal Hospital Infection - Author Affiliations: (1) Queensland University of Technology, Brisbane, Queensland Australia (2) Department of Medicine and Department s of Infectious Diseases and Microbiology, University of Queensland, Mater Misericordiae Hospital, Brisbane, Queensland, Australia. -Funding: not reported	- Design: Retrospective cohort study - Facility/setting: Mater Misericordiae Hospital, 280 beds; Teaching hospital; Hematology/ oncology Ward, ICU, and General wards Brisbane, Queensland Australia - Time Pd: 1/1/1998- 8/1/2002 - Sample: 962 paired peripheral and catheter-drawn cultures from same patient within 120 min of each other. Limited to 1 pair per day. 10 mL blood split evenly between anaerobic and aerobic culture bottles at the bedside. - Comparator: Peripheral Vein Draw (PV) BC drawn from patients within 120 min. of central venous catheter draw. - Skin prep not described - Staff not described - Biases: None noted	- Description: Central venous catheter (CVC); Specimen drawn by ward nursing staff on the haematology/ oncology ward, by resident doctors in the intensive care unit and by trained phlebotomists in general wards. Interlink catheter system cleaned with 70% isopropyl alcohol swabs - Duration: 44 Months -Training: Nursing staff on the hematology/oncology ward, resident doctors in ICU, and trained phlebotomists in the general wards. - Staff/resources: Various staff involved with the BC collections; Bactec 9240 automated BC system -Cost: Not reported	- Description: (1) Blood Culture Contamination Rate (BCCR) – percentage of all blood cultures growing contaminants (2) Specificity* -Recording method: Retrospective chart review and microbiology data.1	- Type of Finding(s): Paired comparison - Finding(s): (1) BCCR:CVC = 125/962 (13.0%); PV = 25/962 (1.3%) (2) Specificity CVC 85% (CI =82% - 87%) PV = 97% (CI= 95%-98%) Effect size: OR = 5.60 (CI, 3.61 – 8.69) - Statistical Significance/Tests: Not conducted for finding, but CI indicates p. <0.05 for specificity - Statistical power is not discussed - Results/conclusion bias: None noted
Quality Rating (10 point maximum): 10 (Good) **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study** **(3 pts maximum): 3**	**Practice** **(2 pts maximum): 2**	**Outcome measures** **(2 pts maximum): 2**	**Results/findings** **(3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Ramsook C (1,2), Childers K (2), Cron SG (3), Nirken M (2) - Year: 2000 - Publication: Infect Control Hosp Epidemiol - Author Affiliations: (1) Eric Williams Medical Sciences Complex (2) Emergency Medicine, Texas Children's Hospital, Baylor College of Medicine, Houston, TX (3) Academic General Pediatrics, Baylor College of Medicine - Funding: not reported	- Design: Observation: IV catheter (IVC) drawn by nurses compared with venipuncture (V) drawn by nurses and phlebotomists - Facility/setting: Texas Children's Hospital; Houston University-affiliated pediatric emergency room Houston, Texas, USA - Time Period: 2/1/1999-7/31/1999 - Sample: Single BC drawn from 2,431 patients - Comparator: venipuncture by nurse or laboratory phlebotomist1 - Bias: None noted	- Description: Single IV catheter (IVC) draw by nurses using three sequential Betadine swabs followed by three sequential alcohol swabs from 1295 patients. - Duration: 6 months - Training: Not discussed - Staff/resources: lab phlebotomist and nurses; aerobic bottles, Betadine swabs, alcohol swabs, saline, Bactec 9240, - Cost: Not reported	- Description: Blood Culture Contamination Rate (BCCR) – percentage of all blood cultures growing contaminants -Recording method: Only aerobic cultures. Review performed during specified time period (6 months) of age, blood site collection, contaminant/ pathogen; practice, and skilled staff performing blood culture.	- Type of Finding(s): non-randomized comparison - Finding(s): (1) BCCR: IVC = 44/1295 (3.4%); V = 22/1084 (2.0%) (2) BCCR: Nurse IVC = 44/1295 (3.4%) (3) BCCR: Nurse V = 5/427 (1.2%) (4) BCCR: Laboratory phlebotomist V = 17/642 (2.6%) Effect Size: OR = 1.70 (CI, 1.01 – 2.85) - Statistical Significance/Tests: - IVC versus V relations not tested - Statistical power is not discussed - Results/conclusion bias: Possible confound by relative skill of person performing draw and by skin preparation.
***Quality Rating** **(10 point maximum): 7 (Fair)** **Effect Size Magnitude : Moderate** **Rating: (Relevance: Direct)**	**Study** **(3 pts maximum): 2** **- Pediatric ED setting may not be generalizable**	**Practice** **(2 pts maximum): 2**	**Outcome measures2** **(2 pts maximum): 1** **- Clinical histories not obtained patients on antibiotics or those immune-compromised not excluded.**	**Results/findings3** **(3 pts maximum): 2** **- Results may be influenced by different preparation systems or age of sample for different practices.**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Norberg, A(1,2), Christopher, N. C.(1,2), Ramundo, M. L. (1,2), Bower, J. R. (1,2), Berman, S. A. (1). **- Year:** 2003 **- Publication:** Journal of American Medical Association **- Author Affiliations:** (1) Divisions of Emergency Medicine and Infectious Diseases, Dept of Pediatrics Children's Hospital Medical Center of Akron, Akron, Ohio (2) Northeastern Ohio Universities College of Medicine, Rootstown, Ohio **- Funding:** not reported	**- Design:** Observational Pre-Post study **- Facility/Setting**: Emergency Department, Children's Hospital Medical Center of Akron, Akron, OH **- Time Period:** 1/1/1998-12/31/1999 **- Sample:** 4,108 Blood culture specimens from emergency department patients <18 yrs. age Venipuncture: 2,000 Cather: 2,108 **- Comparator:** Peripheral intravenous catheter (PIC): blood culture specimens obtained by ED registered nurses through newly inserted peripheral intravenous catheters (1/1/1998-11/19/1998) ; indwelling, vascular catheters excluded **- Study bias:** Pediatric emergency department setting	**- Description**: Venipuncture (V) at a dedicated site distant to the catheter (if required). **- Duration:** 12 months **- Training:** Not discussed **- Staff:** ED registered nurses **- Cost:** Not reported	**- Description:** Blood Culture Contamination Rate (BCCR) – percentage of all blood cultures growing contaminants **-Recording method:** Blinded assessment of pathogenicity or contaminant.[1] - Test for isolates not discussed	**- Type of Finding:** Comparison – Pretest-Posttest **- Finding(s):** (1) BCCR: PIC = 191/2108 (9.1%); V = 56/2000 (2.8%) **Effect Size: OR = 3.46 (CI, 2.55 – 4.69)** **Statistical Significance/Tests:** 2-sample test of proportions *P* <.001 - **Results/Conclusion bias**: None noted
***Quality Rating (10 point maximum): 8(Good)** **Effect Size Magnitude Rating:Substantial** **(Relevance: Direct)**	**Study** **(3 pts maximum): 2** - Pediatric emergency department setting may not be generalizable.	**Practice** **(2 pts maximum): 2**	**Outcome measures** **(2 pts maximum): 1** - Recording method not well described	**Results/findings** **(3 pts maximum): 3**

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Martinez JA (1,2), DesJardin JA (1,2), Aronoff M (1,2), Supran S(1,2), Nasraway SA (1,2), Snydman DR (1,2). **- Year:** 2002 **- Publication:** Critical Care Med **- Author Affiliations:** (1) Departments of Medicine and Surgery, New England Medical Center (2) Tufts University School of Medicine **- Funding:** Supported in part by National Research Service Award T32-AI07329 from NIH	**- Design:** Retrospective cohort study **- Facility/Setting:** Surgical and cardiothoracic ICU, New England Medical Center; Academic Medical Center; 100-300 beds; Boston, MA, USA **- Time Period:** 11/1/1994- 8 1/1997 **-Sample**: 271 critically ill surgical patients in whom samples for paired cultures were drawn by a critical care nurse within 4 hours through a central venous or arterial catheter (C) and peripheral venipuncture (PV) = 490 paired BC. **- *Comparator:** BC drawn through a peripheral venipuncture by critical care nurses; skin disinfected with povidone-iodine. **- Study Bias:** None noted	**- Description:** Central Venous Catheter or Arterial Catheter. Catheter ports or stopcocks were disinfected with either povidine-iodine or 75% isopropyl alcohol **- Duration:** 33 months **- Training:** Not discussed. **- Staff/resources:** Critical care nurses; **-Cost:** Not reported	**- Description:** (1) Blood Culture Contamination Rate (BCCR) – percentage of all blood cultures growing contaminants (2) PPV* (3) Specificity* **-Recording method:** Two physicians blinded to BC source classified paired cultures with at least one positive result as true bacteremia (or fungemia) or contamination.[1] -Each undetected contaminant counted as a separate observation.[2] - A 10-mL blood sample inoculated in aerobic and anaerobic media and processed by using ESP 384 Blood Culture system.	**-Type of Finding(s):** Paired Comparison - Descriptive (one group) **- Findings:** (1) BCCR:C = 20/499 (4.0%); PV = 8/499 (1.6%) (2) PPV: CVC= 63% (CI= 51% - 76%); PV = 78% (CI=64% - 91%) (3) Specificity C 95% (CI =94% - 97%) PV = 98% (CI= 97%-99%) **Effect Size: 2.57 (CI, 1.13 – 5.89)** **- Statistical Significance/Tests:** - Authors use bootstrapping to adjust for potential clustering around patient, hospital admission, and BC since they included multiple blood culture pairs from the same patient.[3] - Statistical power is not discussed **- Results/conclusions biases:** Discrepancy in number of paired cultures 490 in sample description 499 in results[4]
Quality Rating (10 point maximum): 7 (Fair) **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study** **(3 pts maximum): 3**	**Practice** **(2 pts maximum): 2**	**Outcome measures** **(2 pts maximum): 1** - Recording method provides multiple counts for dependent samples	**Results/findings** **(3 pts maximum): 1** **-** Discrepancy in number of paired cultures 490 in sample description 499 in results[4]

Laboratory Medicine Best Practices

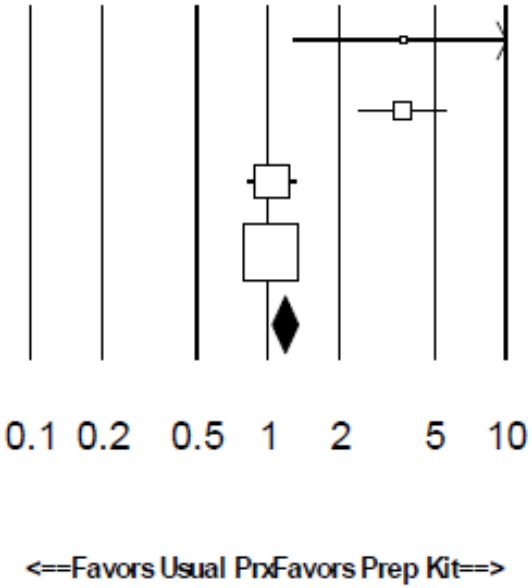
Consensus Ratings 2009

PRE-PACKAGED PREP KITS VS. NO PRE-PACKAGED PREP KITS

Practice: Pre-packaged Prep Kits	Study Quality Rating						Effect Size Rating	Overall Consistency	Overall Strength of Body of Evidence
	Study	Practice	Measures	Results	Total	Rating			
Trautner et al. 2002	2	2	1	2	7	Fair	Substantial	Yes	1 Study = fair/substantial 3 studies = good/minimal High
Wilson et al. 2000	2	1	2	3	8	Good	Minimal/None		
McLellan et al. 2008	2	2	2	2	8	Good	Minimal/None		
Weinbaum et al. 1997	2	2	2	3	9	Good	Minimal/None		

PRE-PACKAGED PREP KITS VS. NO PRE-PACKAGED PREP KITS

ModelStudy name	Subgroup within study	Comparison	Odds ratio and 95% CI
Trautner 2002	Either	Prep Kit v usual prx	
Weinbaum 1997	Combined	Combined	
McLellan 2008	Combined	Combined	
Wilson 2000	Combined	Combined	
Fixed			



**=”Good” Study Quality rating
*=”Fair” Quality rating
Boxes proportionate to study size

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
- Trautner W [1], Jill EC [1, 2], Rabih O D[1,3] - 2002 - Infection Control and Hospital Epidemiology [1] Dept of Medicine, Infectious Diseases Section, Baylor College of Medicine, Houston, Texas [2] Dept of Pathology, Veterans Affairs Medical Center and Baylor College of Medicine, Houston, Texas. [3] Dept of Physical Medicine and Rehabilitation, Center for Prostheses and Infection. **Funding:** Houston Dept of Veterans Affairs and Medi-Flex Hospital Products	- **Design:** Prospective single-blind clinical trial in which each patient received both interventions; both interventions compared with external referent group [1] - **Facility/Setting** VA Medical Center, Houston, Texas; Tertiary-care teaching hospital **Time Period**: 11/12/2000- 5/31/2001 **- Sample:** Paired blood culture sets collected from two venipuncture sites in each of 215 patients, **Comparator:** 383 BC paired specimens (2 venipuncture sites) using povidone iodine swabs or povidone iodine poured onto a cloth swab. BC specimens collected during 1st 6 weeks of the study from patients on the same wards as study patients. **Study bias:** Authors cite 2 potential sources: - "Sicker" patients, requiring greater skill in venipuncture, less likely to be selected to participate by house staff. - House staff more skilled in phlebotomy technique more likely to participate in study.	- **Description**: Paired specimens collected from two venipuncture sites in each pt., using commercial prep kits. [3] **Duration:** 1/1/1998- 12/31/1999 **- Training:** Not discussed **- Staff:** House staff (medical students/residents) and other clinical staff [4] **- Cost:** Not reported	**Description:** Blood Culture Contamination Rate (BCCR) – number and proportion of blood cultures growing contaminant organisms [2] **-Recording method:** Practice: Monitoring of BCs by researcher blinded to collection method Comparator: Not specified.	**Type of finding**: Comparison – RCT with external comparator for both interventions **- Findings/Effect Size:** BCCR: (1) BCCR for venipuncture specimens collected w/o prep kits 25/383 (6.5%) vs. 4/215 (1.9%) for specimens w/ prep kits. (2) Difference in BCCR for the 2 prep kits not stat signif: 0.5% (1/215) with chlorhexidine gluconate kit and 1.4% (3/215) with tincture of iodine plus isopropyl alcohol **Effect size**: OR =3.68 (CI,1.26 – 10.73) **Statistical Significance/Tests:** (1) Fisher exact test, p = .01 [5] (2) McNemar test of dependent proportions, p = .62 Power to detect 4% difference: 0.45 **Results/Conclusion bias**: Staff were aware that contamination rates were being monitored and may have been more careful to follow venipuncture protocol in study patients than in non-participants. Venipuncture technique was NOT directly observed by investigators.

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	**Study** - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	**Practice** - Description - Duration - Training - Staff/Other Resources - Cost	**Outcome Measures** - Description (s) - Recording method	**Results/Findings** - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Quality Rating (10 point maximum): 7 Fair **Effect Size Magnitude Rating: Substantial** **(Relevance: Direct)**	**Study (3 pts maximum): 2** - Primary study focus comparison of 2 interventions - Potential biases	**Practice (2 pts maximum): 2**	**Outcome measures (2 pts maximum): 1** - Recording method for comparator not described	**Results/findings (3 pts maximum): 2** Bias; statistical test
Wilson ML[1], Weinstein MP[2], Mirret S[3], Reimer LG[5,6], Fernando C[5], Meredith FT[3], Reller LB[4]. - 2000 -J Clin Microbiol. -[1] Dept of Pathology and Lab Services, Denver Health Medical Center & U Colorado Schl Med, Denver [2] Clinical Microbiology Lab, Depts. of Pathology & Medicine, Robt Wood Johnson Med Schl, New Brunswick, NJ [3] Clinical Microbiol Lab, Duke U Medl Ctr, Durham, NC [4] Depts. of Pathology & Medicine, Duke U Schl Med, Durham, NC [5] Clin Micro Lab, Salt Lake City VA Med Ctr, Salt Lake City, UT, [6] Dept of Pathology, U Utah Schl Med, Salt Lake City, UT - Funding: Supported in part by Medi Flex (Overland Park, KS)	Design: Non-randomized multi-center comparison study. Facility/Setting: Academic medical center (Robert Wood Johnson University Hospital (RWJUH), Duke University Medical Center (DUMC) and Denver Health Medical Center (DHMC)), and VA/Military/Federal hospital (Salt Lake Veterans Affairs Medical Center (SLVAMC)). Time period: Duration, start/end dates NR Population/Sample: 12,367 blood samples. 6,362 samples collected after disinfection with conventional alcohol pledgets and 6005 collected following disinfection with prep kits.1 Specimens collected via venous catheter excluded from analysis. Comparator: Blood culture specimens collected after antisepsis with conventional isopropyl alcohol pledgets. In each institution, use of prep kits and conventional practice alternated by month. -Biases: none noted	Practice Description: Prep Kit (70% isopropyl alcohol & 2% iodine tincture on separate sterile applicators). Duration: Start/end dates NR. Use of prepkit alternated by month with use of conventional pledgets. Order of use not described. Training: Four sites provided instructions. (RWJUH)- written instructions and verbally (via in-service training). (SLVAMC and DUMC) written instructions (DHMC), kit used for routine skin disinfectant, phlebotomy teams were given verbal instructions via in-service training. - Staff: House staff physicians/medical students except at DHMC, where phlebotomy teams performed venipuncture. - Cost: NR	Outcome measure: Blood culture contamination rate (BCCR)—proportion of all blood cultures growing contaminant organisms.2,3 Recording method: Isolates categorized as clinically important, contaminants, or of indeterminate clinical significance, using std clinical microbiology methods.	Type of Finding: Comparison Findings: Overall contamination rates: • Conventional Pledgets - 5.5% (351 of 6,362 blood cultures; range, 3.7 to 8.1%). • Prep kits: 5.5% (328 of 6,005 blood cultures; range 3.5 to 7.5%). Effect Size: Site ORwtd 95% ci DUMC .1.03 .0.81 – 1.31 RWJUMC .1.08 0.85 -1.39 DHMC .91 0.62 – 1.34 SLVAMC . 1.08 .42 – 2.75 Overall 1.03 0.83 – 1.13 Stat signif/Tests: X2, no significant differences among study sites, no difference in contamination rates between prep kit and conventional pledget skin disinfection Statistical power: Not discussed by authors. Post facto estimate – Power to detect 1.5% overall difference in proportions > .90 Results/conclusion biases: none

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Total Rating (10 pts maximum): 8 Good Effect Size Magnitude Rating: Minimal (Relevance: Direct)	**Study (3 pts maximum): 2 - Duration data collection, order of - Practice/comparator use not stated.**	**Practice:(2 pts maximum): 1 - Variation in implementation of practice among institutions**	**Outcome measures: (2pts maximum): 2**	**Results/findings (3 pts maximum): 3**
Wilson ML[1], Weinstein MP[2], Mirret S[3], Reimer LG[5,6], Fernando C[5], Meredith FT[3], Reller LB[4]. - 2000 -J Clin Microbiol. -[1] Dept of Pathology and Lab Services, Denver Health Medical Center & U Colorado Schl Med, Denver [2] Clinical Microbiology Lab, Depts. of Pathology & Medicine, Robt Wood Johnson Med Schl, New Brunswick, NJ [3] Clinical Microbiol Lab, Duke U Medl Ctr, Durham, NC [4] Depts. of Pathology & Medicine, Duke U Schl Med, Durham, NC [5] Clin Micro Lab, Salt Lake City VA Med Ctr, Salt Lake City, UT, [6] Dept of Pathology, U Utah Schl Med, Salt Lake City, UT - Funding: Supported in part by Medi Flex (Overland Park, KS)	Design: Non-randomized multi-center comparison study. Facility/Setting: Academic medical center (Robert Wood Johnson University Hospital (RWJUH), Duke University Medical Center (DUMC) and Denver Health Medical Center (DHMC)), and VA/Military/Federal hospital (Salt Lake Veterans Affairs Medical Center (SLVAMC)). Time period: Duration, start/end dates NR Population/Sample: 12,367 blood samples. 6,362 samples collected after disinfection with conventional alcohol pledgets and 6005 collected following disinfection with prep kits.1 Specimens collected via venous catheter excluded from analysis. Comparator: Blood culture specimens collected after antisepsis with conventional isopropyl alcohol pledgets. In each institution, use of prep kits and conventional practice alternated by month. -Biases: none noted	Practice Description: Prep Kit (70% isopropyl alcohol & 2% iodine tincture on separate sterile applicators). Duration: Start/end dates NR. Use of prepkit alternated by month with use of conventional pledgets. Order of use not described. Training: Four sites provided instructions. (RWJUH)- written instructions and verbally (via in-service training). (SLVAMC and DUMC) written instructions (DHMC), kit used for routine skin disinfectant, phlebotomy teams were given verbal instructions via in-service training. - Staff: House staff physicians/medical students except at DHMC, where phlebotomy teams performed venipuncture. - Cost: NR	Outcome measure: Blood culture contamination rate (BCCR)—proportion of all blood cultures growing contaminant organisms.2,3 Recording method: Isolates categorized as clinically important, contaminants, or of indeterminate clinical significance, using std clinical microbiology methods.	Type of Finding: Comparison Findings: Overall contamination rates: • Conventional Pledgets - 5.5% (351 of 6,362 blood cultures; range, 3.7 to 8.1%). • Prep kits: 5.5% (328 of 6,005 blood cultures; range 3.5 to 7.5%). Effect Size: Site ORwtd 95% ci DUMC .1.03 .0.81 – 1.31 RWJUMC .1.08 0.85 -1.39 DHMC .91 0.62 – 1.34 SLVAMC . 1.08 .42 – 2.75 Overall 1.03 0.83 – 1.13 Stat signif/Tests: X2, no significant differences among study sites, no difference in contamination rates between prep kit and conventional pledget skin disinfection Statistical power: Not discussed by authors. Post facto estimate – Power to detect 1.5% overall difference in proportions > .90 Results/conclusion biases: none

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
Total Rating (10 pts maximum): 8 **Good** **Effect Size Magnitude Rating: Minimal (Relevance: Direct)**	**Study (3 pts maximum): 2** - Duration data collection, order of - Practice/comparator use not stated.	**Practice:(2 pts maximum): 1** - Variation in implementation of practice among institutions	**Outcome measures: (2pts maximum): 2**	**Results/findings (3 pts maximum): 3**
Weinbaum FI [1,2], Lavie S [3], Danek M [2], Sixsmith D[4], Heinrich GF[5], Mills SS[5]. J Clin Microbiol 1997 Affiliations: The New York Hospital Medical Center of Queens, Departments of [1]Surgery, [2] Quality Management, [3]Pathology, [4] Emergency Medicine, and [5] Administration. Funding: not reported	Design: Nonrandomized prospective intervention trial. Facility/Setting: General adult medical and surgical wards of a community teaching hospital, Flushing, Queens, New York. Time Period: 1 year, dates not specified. Sample: A total of 2562 blood culture specimens collected by a dedicated phlebotomy team and house staff (interns & residents) on two adult in-patient general medical and surgical care units. Comparator: 1. Blood specimens collected by house staff without use of commercial prep kit (N=599) during an initial 3-month interval on Unit A and 2 months on Unit B. 2. Specimens collected by phlebotomy	Description: Two practices were introduced at the same time—a dedicated phlebotomy team and use of a commercial blood culture prep kit with isopropanol and tincture of iodine. Prep kits were initially available only to the phlebotomy team. After the phlebotomy team was withdrawn from one unit, prep kits were provided to the house staff who collected blood culture specimens. Duration: Dedicated phlebotomy team with prep kit: 6 mos. House staff with prep kit: 3 months. Training: not discussed Staff: Three FTE phlebotomists; house staff. Cost: Incremental cost of using the commercial prep kit, rather than alcohol and povidone iodine swabs not	Description: Blood culture contamination rate (BCCR). Recording method: Blood cultures were evaluated by using a BACTEC NR-660 system. Positive cultures were considered contaminated if they grew common skin flora. 2	Type of finding: Comparison between independent groups Findings3: BCRR (se) House staff + prep kit: 4.8 % (.016) House staff – prep kit 8.4%(.014) DPT + prep kit 1.1 % (.004) Effect size4: OR 95% ci DPT + PK v HS – PK 6.25 (.3.58-10.91) HS + PK v HS – PK1.38 (.0.68 – 2.81) (DPT = Dedicated phlebotomy team; PK = prep kit; HS = house staff) Stat tests, significance, & power: Test: Mantel Haenszel X2, $p < 0.001$ for comparisons between house staff without prep kit vs. dedicated phlebotomy team with prep kit. For house staff with prep kit vs house staff without prep kit, $p = .173$. Power not addressed by authors.

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
	team with commercial prep kit (N=1755). Biases: Limitations: Contamination rate for phlebotomy teams that did not use prep kit not collected—improvement in BCCR attributable to prep kit, independent of phlebotomy team, not evaluable.	estimated. 1		Results/Conclusion biases: none
Total Rating(10 pts maximum): 9 Good Effect Size Magnitude: Minimal (Relevance: Direct)	**Study: (3 pt maximum): 2 .- Study limitations**	**Practice: (2 pt maximum): 2**	**Outcome measure(2pt max): 2**	**Results/findings(3 pt max): 3**
McLellan, E; Townsend R, and Parsons HK. -2008 - Journal of Infection - Dept of Microbiology, Northern General Hospital, Sheffield Teaching Hospitals NHS Foundation Trust - Funding: Enturia Limited	Design: Prospective observational intervention study Facility/Setting: Northern General Hospital , Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, South Yorkshire, UK Academic Medical Center Time Period: Two intervention periods— 1st: Aug-Oct, 2007 compared with 2 comparator periods: 1. Aug – Oct, 2006 2. May – Jul 2007 2nd: Dec, 2007 – Feb 2008, one	Description: Blood specimens collected via venipuncture primarily by doctor support workers (DSW). Junior doctors (house staff) obtain blood specimens when DSWs are not available or unable to obtain a specimen. The commercial prep kit, comprising an applicator impregnated with 2% alcoholic chlorhexidine and 2 alcohol swabs to disinfect culture bottle seals, was used on both MAU during 1st	Description: Blood culture contamination rate (BCCR) (by comparison of several time frames) Recording method: All positive blood cultures were evaluated as true positives or contaminants by a medical microbiologist. Positive cultures identified by routine analysis (BACTEC 9240). Positive cultures Gram-stained and subcultured to appropriate	Type of finding: Non-randomized comparison Findings: See Table 1 1 (1) Intervention vs previous year: BCCR 6.6% (20/304) on MAU1 compared with 17.3% during previous year ($p<0.001$) and on MAU2 8.5% (21/248) compared with 13.5% ($p=0.11$) . (2)Intervention compared to immediately preceding 3 months (May-to July 2007): BCCR for MAU1 and MAU2 were 8.7% (P=0.38) and 9.2% (P=0.87), respectively. (3) December 2007- February 2008, prep kit reintroduced only on MAU1;

Bibliographic Information - Author (s) - Yr Published/Submitted - Publication - Author Affiliations - Funding	Study - Design - Facility/Setting - Time Period - Population/Sample - Comparator - Study bias	Practice - Description - Duration - Training - Staff/Other Resources - Cost	Outcome Measures - Description (s) - Recording method	Results/Findings - Type of Findings - Findings/Effect Size - Stat. Significance/Test(s) - Results/Conclusion Bias
	medical admissions unit (MAU1) in hospital was compared to another (MAU2). Sample: 1st intervention period: 552 BCs 2nd int period: 610 BCs. See Table 1. 1 Comparator: Blood culture contamination rates (BCCR) for the corresponding MAU during same period in 2006 (525 BCs) and 3 months immediately preceding 1st intervention period (563 BCs). Biases: None noted	intervention period, but only on MAU1 during 2nd period. Duration: Aug-Oct 2007; Dec 2007- Feb 2008 - Training: Kit manufacturer provided training to DSWs, house staff, and nursing staff. Infection control staff provided training on BC site preparation and consequences of BC contamination. - Staff: No changes in staffing required to implement practice. Units staffed by 10 Jr doctors, one DSW per day shift and a single DSW to cover both wards during the night. - Cost: Unit cost for prep kit: £0.50; unit cost for 70% isopropyl alcohol wipes: £0.02.	solid agar; isolates were identified by "multipoint methods" and bench-level diagnostics. The same routine methods were in use during comparator periods.	contamination rate = 8.5% (28/330). Not significantly different from either 1st intervention period (Aug-Oct 2007) or pre-implementation period (May-Jul 2007); (P=0.37 and P=1.0, respectively). (4) BCRR on MAU2 Dec 2007 – Feb 2008: 6.4% (18/22), non-significant reduction (P = 0.41) compared with intervention period, despite the absence of prep kit. Effect size3: OR 95% ci MAU1 1.16 ..72-1.87 MAU2 0.89 0..51 – 1.56 Overall 1.03 .73 – 1.47 Stat tests & Power: X2 Statistical power is not discussed Results/Conclusion biases: Authors present evidence of confounding by a time-related factor. Drop in contamination rate between 2006 and 2007 appears unrelated to use of prep kits.
Total Rating(10 pts maximum): 8 Good Effect Size Rating: Minimal (Relevance: Direct)	**Study: (3 pt maximum): 2 - 1Comparator period non-representative**	**Practice: (2 pt maximum): 2**	**Outcome measure(2pt max): 2**	**Results/findings(3 pt max): 2 - Confounding, limited statistical analysis**

APPENDIX F. GUIDE TO RATING STUDY QUALITY

Individual study quality ratings are based on four dimensions of study quality:

- Study
- Practice
- Outcome Measure(s)
- Findings/Results

The principles and guidelines for making judgments along these four dimensions are outlined below.

The objective for rating individual study quality is to judge whether sufficient evidence is available concerning a practice's effectiveness to support a recommendation of "best practice" (that is, a practice likely to be effective in improving one or more outcomes of interest in comparison to other commonly used practices).

The rating methods are designed to be inclusive, so we can make best use of the limited data available, which may commonly be in the context of a quality improvement effort. Our methods for rating study quality do not penalize studies for not using a randomized design or for not being published in a peer reviewed journal. This approach acknowledges that many practices in laboratory medicine do not lend themselves to evaluation by traditional research designs, and much useful evidence may be obtained though our network affiliates whose priority on service delivery takes precedence over publishing.

Evaluating Study Quality

The four study quality dimensions are rated separately, with a rating score assigned up to the maximum for a given dimension. The rating scores for each dimension are added to reach a single summary score reflecting overall study quality. A total of 10 points are available for each study, with points subtracted from the maximum point total according to the guidance below. Anytime points are deducted from a study, a justification for the deduction is recorded and included in the evidence summary. In this scheme, a rating of zero in any one of the four categories is sufficient to exclude a study from further consideration as evidence for a "best practice" recommendation.

Dimension 1. Study (3 points maximum)

Assess the likely generalizability of the results by evaluating:

- Study setting
- Sample characteristics (representativeness sufficient for practice)
- Potential study biases (study design, time period/duration and sample selection methods)

Criteria for point deduction

Rate Facility Description / Study Setting: From the drop-down list provided, rate the uniqueness of the study facility and setting with respect to generalization. Generalizability is a judgment on the likelihood that the results obtained would achieved in other facilities or settings.

1. Deduct 0 points. Results are likely to generalize to other facilities and settings.

2. Deduct 1 point if the study location is sufficiently distinctive that the results obtained through that setting may not be generalizable to other settings.

3. Deduct 2 points if the study location is sufficiently distinctive that the results obtained through that setting are unlikely to be generalizable to other settings

4. Deduct 3 points (Score 0) if it is clear that the setting or situation is unique such that the results cannot generalize to other settings.

Rate Study Design / time period / population: From the drop-down list provided, rate the uniqueness of the study design, time period, or sample with respect to representativeness. Representativeness is a judgment on whether the results obtained through the study design are likely representative of the results of the practice.

1. Deduct 0 points. The sample / time period / and population is sufficient to be representative of the results of the practice.

2. Deduct 1 point if the sample (either subjects or tests) may not be representative of the likely results of the practice with respect to how the sample was obtained or identified

3. Deduct 2 points if the sample (either subjects or tests) are probably unrepresentative of the results of the practice with respect to how the sample was obtained or identified.

4. Deduct 3 points if the sample is sufficiently unrepresentative based on how it was obtained/identified to clearly nullify the generalizability of the results.

Rate Potential Study Bias: From the drop-down list provided, rate the extent to which you believe that the study design, period of measurement, and/or sample selection introduced bias to the results. Bias is systematic error that either reliably enhances *or* suppresses a finding. It is different from error, which reduces the quality of measurement, but not in any systematic fashion (can produce findings that are both larger and smaller than the true score).

1. Deduct 0 points

2. Deduct 1 point if the study design, time period and sample selection methods may introduce a study bias that would substantially affect results (i.e., may produce study results interpreted as inconsistent with the true results)

3. Deduct 2 points if the study design, time period and sample selection methods are likely to introduce a study bias that would substantially affect results (i.e., would likely produce study results interpreted as inconsistent with the true results)

4. Deduct 3 points if there is reason to believe that the study characteristics can not produce results representative of the practice

Total Points Deducted – Study: Using the information summarized by the three previous Study Rating Items (Items Rate Facility Description / Study Setting; Rate Study Design / time period / population; and Rate Potential Study Bias respectively), record in this field the maximum number of points deducted in any of the previous three Study rating fields. For example, if rows 13 and 23 are each zero, but row 32 equals 3 points deducted, then the total points deducted would be 3. If each of the three previous ratings are 2, then 2 points would be selected for this item.

3 points deducted	2 points deducted	1 point deducted	0 points deducted

Dimension 2. Practice (2 points maximum)

Assess the description of the practice and its adequacy.

Criteria for point deduction

Rate Practice Description: The practice should be well enough described to meaningfully distinguish it from alternative practices and provide a clear understanding of its requirements and characteristics (does not require that the description be exhaustive or support exact replication).

Deduct 0 points: The practice is well described

Deduct 1 point if the practice and its basic characteristics are not sufficiently identified.

Deduct 2 points (Score 0) points if the practice and its basic characteristics cannot be clearly identified.

Rate Adequacy of Practice Description: Ideally, seven components of practice description would be addressed: a) content, b) implementation, c) population / setting, d) training, e) requirements, f) cost, and g) staff responsible and implementing. However, detailed information on all components is not necessary to evaluate the effectiveness and feasibility of implementing a practice, and is typically not provided (e.g., cost and training).

Deduct 0 points: Based on the practice description, the practice could be adopted in other settings.

Deduct 1 point if an important aspect/component that is likely to critically affect implementation of the practice is not well described.

Total Points Deducted – Practice: Using the information summarized by the two previous Practice Rating Items (Rows 16 and 18; Items Rate Practice Description and Rate Adequacy of Practice Description respectively), record in this field the maximum number of points deducted in any of the two previous Practice rating fields. The options for this item are:

0 points deducted 1 point deducted 2 points deducted

Dimension 3. Outcome Measure (2 points maximum)

Outcome measures capture the result of implementing a practice. Evaluation criteria reflect their face validity for capturing the outcome(s) of interest, and whether the methods used to record results provide an incomplete or inaccurate record of the impact of a practice.

Most studies use multiple outcome measures. Their evaluation should concentrate on measures that most directly address the review question, which relates to health care quality (Institute of Medicine domains: safe, timely, effective, patient-centered, efficient, and equitable), and may ignore secondary measures, especially those gauging implementation feasibility.

Criteria for point deduction

Rate Face Validity of Outcome Measure: Face validity asks if the measure likely captures the construct it is intended to measure. Deduct points if the measure as specified/defined and data collected is possibly or likely flawed and/or if the measure does not capture the essence of what it was supposed to measure (e.g., timeliness). Use the following criteria in making your judgment

1. Deduct 0 points/no deduction

2. Deduct 1 point if:

(1) Measure does not capture well the outcome being estimated OR

(2) The measure estimates an outcome that is only modestly related to the evidence review question (e.g., provider satisfaction, compared with change in a patient outcome or an error rate)

3. Deduct 2 points if the measure is confounded by:

(1)The practice itself (that is, the outcome is a direct result of the practice which was not available or applicable to both the comparison practice and the tested practice) OR

(2) The measure is confounded by the context in which the practice has been implemented (that is, the outcome is unlikely to be clearly attributed to the practice).

Rate Recording Method: The method for recording or documenting practice results should be reliable and accurate. As mentioned above, recording methods may be subject to bias or fraud (intentional and otherwise). Reliability refers to the ability of a measure to return the same result each time, given the same underlying phenomena. Accuracy refers to the precision of the measurement tool. Note that a tool can be accurate, but not reliable (e.g., the adage “measure twice, cut once” observes that while the tape is accurate, the observation of the result is not reliable). Use the following scale to rate the quality of the recording method described.

1. Deduct 0 points

2. Deduct 1 point if the recording method is not described OR does not accurately capture all instances of the outcome

3. Deduct 2 points if the method of recording the outcome is unreliable.

Describe Recording Method Rating: If points were deducted for the recording method of the outcome measure, please provide a rationale for the deduction.

Total Points Deducted – Measurement: Using the information summarized by the two previous Measurement Rating Items (Rows 7 and 11; Items Rate Face Validity of Outcome Measure and Rate recording method respectively), record in this field the maximum number of points deducted in any of the two previous Measures rating fields. The options for this item are:

2 points deducted 1 point deducted 0 points deducted

Dimension 4. Results/Findings (3 points maximum)

Results are affected by each of the previous three dimensions of quality. With this dimension, a narrow set of criteria specific to the result are evaluated relating to (1) sample sufficiency, (2) appropriateness of statistical analysis and, (3) uncontrolled deviations along with results/conclusions bias.

Criteria for point deduction

Rate Sample Sufficiency: Many of the outcomes of interest are rare events. If too few observations are obtained or if the measurement period is insufficient to capture these events the measure may provide an inaccurate representation of the effect of the practice. Even among more common events, there may also be considerable variation in the number or rate of events over time. The period of measurement should be sufficiently long to allow robust estimates of the impact of the practice.

Deduct 0 points

Deduct 1 point if:

(1) Measurement period insufficient to allow robust estimate of practice impact) OR

(2) Statistical power not discussed AND the sample may be too small to allow a robust estimate)

Deduct 2 points if:

(1) The number of subjects and/or measurement period not reported)

(2) The measurement period is likely insufficient for a robust estimate of practice impact) OR

(3) Statistical power is not discussed AND the sample is likely too small to allow a robust estimate of the impact of a practice)

Describe sample sufficiency rating: If points were deducted for the Sample Sufficiency Rating, please provide a rationale for the deduction.

Rate appropriateness of statistical analysis: Determine if the statistical analysis was appropriate for the data.

Deduct 0 points.

Deduct 1 point if:

(1) Data collected during notably different time periods OR

(2) Data do not permit effect size calculation.

Deduct 2 points if:

(1) Results were obtained using different measures or recording practices for two groups

(2) Inappropriate statistical analysis AND insufficient data to allow or verify calculation of an effect size)

Rate Uncontrolled Deviations and Results/Conclusions biases: Despite the best efforts of researchers, studies are sometimes contaminated by unforeseen events or circumstances. Record here the extent to which these may apply to the findings being recorded.

Deduct 0 points

Deduct 2 points if:

(1) Results not due to practice; likely related to different practice(s), OR

(2) Unexplained attrition >70% OR

(3) Uncontrolled differential attrition OR

(4) Results/conclusion not supported by or at odds with the work done/presented

Describe bias rating: If the findings were influenced by Uncontrolled Deviations or Results/Conclusions biases, describe the uncontrolled deviations and results/conclusions biases in this field.

Total Points Deducted – Findings: Using the information summarized by the three previous Findings Rating Items (Rows 15, 17, and 19; Items Rate Sample Sufficiency, Rate appropriateness of statistical analysis, and Rate Uncontrolled Deviations and Results/Conclusions biases respectively), record in this field the maximum number of points deducted in any of the three previous Findings rating fields. The options for this item are:

0 points deducted	1 point deducted	2 points deducted	3 points deducted

Quality Dimensions	Maximum Points	Rating Criteria	Deduct 1 Point if:	Deduct 2 Points if:	Score Zero if:
Study	**3**	Facility Description	The study location is sufficiently distinctive that the results obtained through that setting may not be generalizable to other settings.	The study location is sufficiently distinctive that the results obtained through that setting are unlikely to be generalizable to other settings	It is clear that the setting or situation is unique such that the results cannot generalize to other settings
		Study design/study time period/ patient population	The sample (either subjects or tests) may not be representative of the results of the practice with respect to how the sample was obtained or identified AND the non-representativeness suggests that the results may not be generalizable.	The sample (either subjects or tests) are unlikely to be representative of the results of the practice with respect to how the sample was obtained or identified AND the non-representativeness suggests that the results are unlikely to be generalizable.	The sample is sufficiently unrepresentative based on how it was obtained/identified to clearly nullify the generalizability of the results.
		Potential study bias	The study design, time period and sample selection methods may introduce a study bias that would substantially affect results (i.e., may produce study results interpreted as inconsistent with the true results)	The study design, time period and sample selection methods are likely to introduce a study bias that would substantially affect results (i.e., would likely produce study results interpreted as inconsistent with the true results)	There is reason to believe that the study characteristics can not produce results representative of the practice
Practice	**2**	Description of practice	The practice and its basic characteristics are not sufficiently identified.	N/A	The practice and its basic characteristics can not be clearly identified.
		Adequacy of practice description	An important aspect/component that is likely to critically affect implementation of the practice is not well described.	N/A	N/A

Outcome Measures	**2**	Face validity	The 'best' measure from a study: - Does not capture well the outcome being estimated **OR** - Estimates an outcome that is only modestly related to health care quality or patient safety (e.g., provider satisfaction, compared with change in a patient outcome or an error rate)	N/A	The 'best' study measure: - Is confounded by the practice itself (outcome is a direct result of the practice; not available to the comparison) **OR** - Is confounded by the context in which the practice has been implemented (outcome is unlikely to be clearly attributed to the practice). **OR** - Does not have the potential to contribute to health care quality or patient safety
		Recording Method	Method(s) of recording: - Not described **OR** - Does not accurately capture all instances of the outcome	N/A	Method of recording the outcome is unreliable.
Results / Findings	**3**	Sample Sufficiency	The measurement period may be insufficient to allow a robust estimate of the impact of a practice. **OR** Statistical power is not discussed AND the sample may be too small to allow a robust estimate of the impact of a practice	Number of subjects not reported **OR** Measurement period not reported **OR** Measurement period likely insufficient for a robust estimate of the impact of a practice **OR** Statistical power is not discussed AND the sample is likely too small for a robust estimate of the impact of a practice	N/A

		Appropriateness of statistical analysis	Compares two practices and their estimates are based on data collected during notably different time periods **OR** Does not provide data sufficient to allow/verify calculation of an effect size	Different measures or different recording practices are used when comparing the results of two practices **OR** An inappropriate statistical analysis *and* insufficient data to allow/verify calculation of an effect size	N/A
		Uncontrolled Deviations and Results/conclusion bias	N/A	Results/effect size reported not clearly attributable to practice being evaluated, but instead are likely related to significantly different practice(s) **OR** There is unexplained attrition > 70% or the study uses a randomized design and there is differential attrition not controlled by analysis **OR** Results reported and/or conclusions are not representative of or supported by the work that was done (e.g., additional relevant findings are mentioned, but not reported and/or not incorporated in the conclusions; results conflict or are at odds with conclusions).	N/A
Overall Study Quality	**10**				

APPENDIX G. EFFECT SIZE RATING GUIDANCE

Background

Reports about laboratory medicine practices, whether published or not, are often descriptive and opportunistic, taking advantage of data developed for in-house monitoring and practice improvement. Less attention is paid to research methods than might be expected. In evaluating reported findings, reviewers should pay particular attention to the outcome measures used and methods of measurement. Many other issues that are typically evaluated when assessing study quality are either not relevant to observational study designs or reported in a way that is evaluable with any degree of sophistication. We urge reviewers to give this evidence the benefit of the doubt and as such, not hold it to a standard to which it does not aspire.

Effect Size Rating Categories

Reviewers will rate the reported effect size for a given study in one of three categories:

- Substantial
- Substantial
- None/Minimal
- Adverse

These ratings are specific to topic areas, requiring each Expert Panel to specify value ranges for each category for the relevant outcome measures. Note that studies may provide a range of results on various outcome measures. When rating effect sizes in support of best practice recommendations, please choose the highest rating supported by study evidence. That is, if a practice has multiple effect sizes some of which are “moderate” and some of which are “substantial,” rate the impact of the practice as substantial.

- To receive a “Substantial rating” the study reports a measure that directly answers the review question and has
    - No relevant adverse effects AND
    - At least one good effect large enough to clearly support practice implementation
- To receive a “Moderate rating” the study reports a measure that directly answers the review question and has
    - No relevant adverse effects AND
    - A good effect large enough to favorably support practice implementation
- To receive a “None-minimal rating” the study reports
    - No relevant adverse effects AND
    - An effect size that is similar to zero or slightly positive on a measure that directly answers the review question asked

Considerations for Evaluating Effect Size: Magnitude and Relevance

Evaluating reported effect sizes involves judgments of both the magnitude of the effect observed, and how centrally relevant the measure is to answering the review question. These are two separate judgments, and each needs to be considered when evaluating evidence for contributing to a best practice recommendation.

Judging the magnitude of effect. Little if any of the evidence available for this task will be based on randomized designs. The typical LMBP study uses a pre-post one-group design. That is, the study provides a before-implementation estimate of a previous standard or "comparison" practice on some measure and an after-implementation estimate of the new or "tested" practice on the same measure. The next most common design is a single (post-test only) after-implementation estimate on some measure of interest.

In contrast with controlled research, the comparison practice against which a new practice is tested will likely vary from study to study. This can affect the difference score (the finding) obtained as much as the new practice. When interpreting magnitude of effect, consideration should be given to both halves of the comparison as the finding represents the difference between the two that may not be the impact of the new practice over a common base.

To aid reviewers in judging the magnitude of effect between a new/tested practice and a comparison practice, when possible we calculated two standardized measures as findings for each reported outcome.

(1) Odds Ratio (OR) [7] compares the chance of an event occurring in one group versus another group (e.g., new/post-practice and standard/pre-practice) for dichotomous outcomes (i.e., 2 possible outcomes such as yes/no; error/no error) and has the following interpretation:

- OR > 1: new practice is more successful than the standard practice; the larger the number, the greater the relative success
- OR = 1: new practice is equal to the standard practice,
- OR<1: new practice is less successful than the standard practice; the smaller the number, the worse its relative success

For example, OR = 1.5 means the tested practice is half again as successful as the old practice, OR = 2.0 means the tested practice is twice as successful as the standard practice, and OR 0.66 means the tested practice is two thirds as successful as the standard practice.

(2) Cohen's *d*.[8] (*d*-score) compares an effect size to zero, and has the following interpretation:

- *d*-score > 0: new practice is more successful than standard practice
- *d*-score = 0: no differences between new practice and standard practice
- *d*-score <0: new practice is less successful than standard practice

The further the *d*-score is from zero the more successful the practice is relative to the comparison practice when positive, and the less successful when negative.

Quality of the Outcome Measure for Contributing to a Recommendation. Since the studies considered were not designed to directly address the LMBP review question, not all outcome measures in the findings reported have equal relevance to the LMBP review question.[9] The highest quality measures

---

[7] See Attachment A for more details on how these standardized metrics are calculated.

[8] See Attachment A for more details on how these standardized metrics are calculated.

[9] The Study Quality Rating Guide has reviewers rate the quality of the measure for meeting the author's purpose, is it a reliable and valid measure of their outcome of interest. The Effect Size Rating Guide has reviewers rate the quality of the measure for answering LMBP's review question. A measure can be rated high when estimating study quality, but may rate low with respect to relevance to the review.

for the purpose of a review are those that are most responsive to the LMBP review question. In the case of the patient-specimen identification practices, the review question's objective is accurate matching of patient and specimen.. In the case of critical values communication practices, the review question's objective is timely and accurate reporting of laboratory test critical values to licensed/responsible caregivers.

Proximal outcome measures are of greater value than distal measures. "R*elevant to LMBP review questions*" is not identical to "*improving health outcomes.*" As our Workgroup and Expert Panel members have observed, health outcomes usually result from a number of contributing factors, and while effective laboratory medicine practices are *necessary* contributors, they may not be *sufficient*. The best estimates of the true impact of a practice should be outcomes that are the direct and immediate consequence of the practice.

Without minimizing the importance of assessing patient health and safety outcomes, the practices being studied occur in a continuum of care, so isolating the impact of a change in practice on patient health and safety may be confounded by health care delivery and other potential moderating factors. As a purpose of LMBP evidence reviews is to document the evidence for best practice recommendations, reviewers should give additional weight to outcomes that are a direct result of a practice that directly address the review question, and discount findings that are likely affected by additional aspects of health care delivery or other identifiable factors.

Outcome measures that more directly address the review question are more relevant to its effect size

Also reported are measures that should not be interpreted as an outcome of direct interest (i.e., does not directly address the review question objective), but may contribute to judgments of feasibility as they are within the causal field for the practice (e.g., percentage of patients correctly identified using bar-coding but no specimens involved, percentage of critical values reported using a call center without information related to accuracy or timeliness),. In the language of causal reasoning, they may be necessary, they may be important, but they are not sufficient evidence of practice effectiveness for making a "best practice" recommendation.

Attachment A: Calculating Effect Sizes

1. When dichotomous or binary outcomes were addressed (e.g., success versus failure) and the findings provided rate or percentage information, we calculated an odds-ratio (See Formula 1) that contrasts the two groups in terms of the relative odds of a successful outcome (e.g., the patient either is or is not correctly matched with their specimen).

Table 1: Odds Ratios

	Frequencies		Proportions	
	Success	Failure	Success	Failure
Tested Practice	a	B	$p_a = a/(a + b)$	$p_b = b/(a + b)$
Comparison Practice	c	D	$p_c = c/(c + d)$	$p_d = d/(c + d)$

$$ES_{OR} = \frac{ad}{bc} = \frac{p_a p_d}{p_b p_c} = \frac{p_a / p_b}{p_c / p_d} = \frac{p_a(1-p_c)}{p_c(1-p_a)} \quad \textbf{Formula 1 (Odds Ratio)}$$

Where Pa = a/(a + b), Pc = c/(c + d) and a, b, c, and d are proportions in a four-square cross-tabulation table.

2. For both dichotomous outcomes and for outcomes based on continuous data, we calculated a Cohen's d (See Formula 2), which can be interpreted as the difference between the groups on a standardized scale with a mean of zero (meaning no effect) and a standard deviation of 1 (i.e., a normal distribution). Positive results signify that the finding favors the post-test or 'new' practice) while negative findings favor the comparison practice.

$$ES_d = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{s_p} = \frac{\overline{X}_{G1} - \overline{X}_{G2}}{\sqrt{\frac{(n_{G1}-1)s_{G1}^2 + (n_{G2}-1)s_{G2}^2}{(n_{G1}-1)+(n_{G2}-1)}}} \quad \textbf{Formula 2 (Cohen's d)}$$

Where $\overline{X}_{G1}$ = $\overline{X}_{G1}$ *= the mean for Group 1,* $\overline{X}_{G2}$ = $\overline{X}_{G2}$ *= the mean for Group 2,* S_p = S_p *= the pooled standard deviation of the two groups,* n_{G1} *= number of tests in Group 1,* n_{G2} *= number of tests in Group 2* s_{G1} *= the standard deviation of tests in Group 1, and* s_{G2} *= the standard deviation of tests in Group 2.*

Cohen's-*d* can be approximated from many other reported statistics (e.g., *t-tests, chi2, one-way ANOVA, frequency, and proportion)* and can be converted into multiple indices for interpretation (see Table 2).

Table 2: Cohen's-*d* Expressed in Alternative Indices

Cohen's-*d*	**r**	**% Variance Explained (r^2)**	**U3 (% of T above Xc)**	**BESD C v. T Success Rates**	**BESD C v. T Differential**
0.0	.0	.0	.50	.50 v .50	.0
0.1	.05	.003	54	.47 v .52	.05
0.5	.24	.06	.69	.38 v .62	.24
1.0	.45	.20	84	.27 v .72	.45
1.5	.60	.36	93	.20 v .80	.60
2.0	.71	.50	98	.14 v .85	.71
2.5	.78	.61	99	.11 v .89	.78
3.0	.83	.69	99	.08 v .91	.83

Where r = correlation, T = treatment successful, Xc = comparison median, C = comparison successful; BESD = Binomial Effect Size Display

APPENDIX H. DATA ABSTRACTION CODEBOOK

This coding manual summarizes the rules and guidelines for abstracting evidence for best practices in laboratory medicine. it is appropriate to discuss coding issues to try to reach consensus on coding principles. As a convention, please use NA to indicate "not applicable" and NR to record "not reported" in fields where data entry is not fixed (i.e., not restricted to drop-down options).

If you have questions or desire guidance, do not hesitate to contact Jim Derzon directly (derzonj@Battelle.org; 703-248-1640).

Table 1: References

This table captures all references identified for the study. It provides tracking information for monitoring document status, records how and to which study they are relevant, and includes information on document study eligibility. Note: there may be multiple references (documents) for a single study-sample.

Short Name for Study Citation: Use the format "Topic_Practice_First Author_Year" indicating the Topic, practice, last name of the first author, and year of publication or receipt of unpublished data. Use previously identified and standardized topic and practice names (i.e., do not use ones that have not been pre-approved as appropriate for a specific evidence review).

Topic Area – Practice: Write out the full label for the standardized name of the Topic Area and Practice that the document addresses.

Gives Permission to be Identified (unpublished only):

For unpublished studies indicate whether the author has given permission to be identified. Select "1. Yes" if the author agrees to be identified, "2. No" if they wish to remain anonymous, and "9. Not Applicable" for published studies.

How was document identified?: Indicate how the document was originally identified for this review.

1. Bibliographic literature search
2. Hand search of relevant journals
3. Snowball sample (identified through review or another study bibliography)
4. Response to direct solicitation
5. Identified by key informant
6. LMBP Network

Date eligibility review completed (MM/DD/YY) and Eligibility Reviewer (Initials)

This information provided by review coordinator making assignments.

Select the type of document: Select the option that reflects best how the document was made available for dissemination. Note that technical reports can be either published or not.

Published documents

1. Professional Guidelines
2. Peer-reviewed journal
3. Book / book chapter
4. Technical Report

5. Conference proceeding

Unpublished documents

6. In-house audit or Quality Control Initiative

7. Technical Report

8. Manuscript

9. Conference presentation

10. Other

Describe other document type: If "10 Other" is selected describe the document type here.

Complete Citation or Identifying Information for Unpublished Source. Enter the full citation using standard AMA style. Guidance can be found at http://www.nlm.nih.gov/bsd/uniform_requirements.html. For unpublished studies please identify the source completely as these files are confidential.

Author Affiliations [1], [2], etc. Record the affiliations of all authors. Use the convention Author 1 (1,2), Author 2 (2), Author 3 (1,3). 1 = Institution 1, 2 = institution 2, 3 = Institution 3. If the document has no authors identified, the affiliation of the "First Author" should be the facility or healthcare organization and/or system that provide the study. The affiliation for the "second author" should be the alternate point of contact for the facility or health system that provides the unpublished study.

Funding Source(s): Identify (name) the actual source of funding for the research conducted.

Reviewer (Initials): Select the initials of reviewer responsible for the abstraction. If it is a consensus review, enter the 'lead' reviewer, the reviewer who would be most qualified / comfortable responding to questions that could arise later in the review process.

Other references: Enter any other references used to code this study, or sources referenced in the study that offer significant information related to the study itself, the practice or topic area for background or other considerations such as applicability, feasibility of implementation, costs or other outcomes.

Table 2: Study

This section abstracts information on the study and setting characteristics that may be important for contextualizing the results and also abstracts information on the sources of evidence included in the study. This table leans towards describing observational studies and contains items that are likely to be valuable in describing findings constraints (e.g., "good evidence for this approach, but only for some populations and types of facilities").

Short Name for Study Citation: Topic_Practice_Author_Year: Copy and paste using data recorded in the Biblio field of the same name.

Study reviewer: Copy and paste using data recorded in the Biblio field of the same name.

Name of the facility(ies) or organization(s) where study performed: Enter the name of the facility or group (e.g., Hospital A, College of American Pathologists) that is the source of the data.

Facility/Organization Type: Select from the pull down list the type of facility in which the study was conducted.

Options include:

1. Academic Medical Center

2. Teaching hospital

3. Non-teaching hospital

4. VA/Military/Federal Government Hospital

5. Outpatient Laboratory

6. Physician Office Laboratory

7. Public Health Laboratory
8. Independent / Commercial Laboratory
9. Blood Bank
10. Other (please specify)

Facility/Organization Type - "Other" please specify: If the facility in which the study was conducted is not one of the first five options above, use this text field to describe the facility or facilities in which the study was conducted.

Facility Size: Select from the pull down list the option that best describes the size of the facility in which the study was conducted. Options (for hospitals) include:

1. <100 beds, 2. 100-300 beds 3. >300 beds 4. Not applicable

Total Annual Test Volume: Please enter the option that best corresponds to the annual test volume of the facility in which the study was conducted. Use the "Other" field to provided additional information, if appropriate.

1. <100.000
2. 100,001-500,000
3. 500,001 – 1,000,000
4. > 1,000,000
5. Other
6. Not available

Total Annual Test Volume – "Other" please specify

If total annual test volume standard options do not describe available annual test volume information, or to include information about annual test volume that is not inclusive of all laboratory testing (e.g., specific test volume, anatomic pathology, microbiology, chemistry) use this text field to describe available test volume information.

Facility location - Published Study or Unpublished Study with permission to identify [City, State, Country]: Record the location of the facility in which published study was conducted. Please include the City, State, and Country separated by commas. If multiple locations are represented, please separate locations using a semi-colon.

Facility location - Unpublished Study - No permission to identify [Region, Country]: For unpublished studies, record the region and country of the facility in which the study was conducted. In the United States, states can are defined by their census region and division. Use the region and division designations contained in the parentheses for the 50 United States:

Alabama (R3D6)
Alaska (R4D9)
Arizona (R4D8)
Arkansas (R3D7)
California (R4D9)
Colorado (R4D8)
Connecticut (R1D1)
Delaware (R3D5)
District of Columbia (R3D5)
Florida (R3D5)
Georgia (R3D5)
Hawaii (R4D9)
Idaho (R4D8)
Illinois (R2D3)
Indiana (R2D3)
Iowa (R2D4)
Kansas (R2D4)
Kentucky (R3D6)
Louisiana (R3D7)
Maine (R1D1)
Maryland (R3D5)
Massachusetts (R1D1)
Michigan (R2D3)
Minnesota (R2D4)
Mississippi (R3D6)
Missouri (R2D4)
Montana (R4D8)
Nebraska (R2D4)
Nevada (R4D8)
New Hampshire (R1D1)
New Jersey (R1D2)
New Mexico (R4D8)
New York (R1D2)
North Dakota (R2D4)
North Carolina (R3D5)
Ohio (R2D3)
Oklahoma (R3D7)
Oregon (R4D9)
Pennsylvania (R1D2)
Rhode Island (R1D1)
South Carolina (R3D5)
South Dakota (R2D4)
Tennessee (R3D6)
Texas (R3D7)
Utah (R4D8)
Vermont (R1D1)
Virginia (R3D5)
Washington (R4D9)

West Virginia (R3D5) Wisconsin (R2D3) Wyoming (R4D8)

Setting within facility / organization where practice implemented: Select from the pull down list the option that setting within the facility or organization that best describes where the practice – as tested – was implemented. Options include:

1. Hospital inpatient
2. Hospital outpatient
3. Emergency Department
4. ICU
5. Physician Office
6. Other - please specify

Setting: If "Other", please specify: If the setting in which the practice – as tested – was implemented is not well described by the first five options above, use this field to describe the setting in which the practice – as tested – was implemented.

Rate Facility Description / Study Setting: From the drop-down list provided, rate the uniqueness of the study facility and setting with respect to generalization. Generalizability is a judgment on the likelihood that the results obtained would achieved in other facilities or settings.

1. Deduct 0 points. Results are likely to generalize to other facilities and settings.

2. Deduct 1 point if the study location is sufficiently distinctive that the results obtained through that setting may not be generalizable to other settings.

3. Deduct 2 points if the study location is sufficiently distinctive that the results obtained through that setting are unlikely to be generalizable to other settings

4. Deduct 3 points (Score 0) if it is clear that the setting or situation is unique such that the results cannot generalize to other settings.

Additional Information about Facility Description Study Setting Rating: If you deducted points in the item above, provide a supporting explanation for point deductions.

Study Population: Description of the population (patients, specimens, and /or tests) total number: Provide a rich description of the study population(s) that are included as units of analyses in the report. Be sure to record all described characteristics for both the entire study sample(s) and any breakouts tested. Be sure to include the number of patients and/or number of tests and/or number of samples in total and in each condition, information about the patient type, and any other characteristics of the patient, specimen, and/or tests that are summarized in the study being coded.

Data Collection Start Date (MM//DD/YYYY): Record the month and year that data were first collected for this study. Although the program formats the data as MM/YYYY, data must be entered using the MM/DD/YYYY format. If the day and month are not reported, enter your best estimate of the start of data collection.

Data Collection End Date (MM/DD/YYYY): Record the month and year that data were last collected for this study. Although the program formats the data as MM/YYYY, data must be entered using the MM/DD/YYYY format. If the day and month are not reported, enter your best estimate of the end of data collection.

Describe Data Collection: Many data collection designs are not readily captured using standard start and end dates. Use this field to describe the data collection protocol. Be sure to record dates for each condition. For example: The study used a pre-post design with comparison practice data collected 12/01/1996 to 12/31/1999 and tested practice data collected between 12/01/2007 through 12/01/2009.

Method for selecting tests: From the drop down list provided, select the option that best describes which tests or samples were included in the study.

1. Study includes a census of tests (e.g., all tests within a given time period), and the time and/or number of tests is sufficient to be representative

2. Study includes a randomized or stratified random sample of tests

3. Study used a quasi-random procedure to identify the sample of tests (e.g., every third test was included in the study)

4. Study used a non-random procedure to identify the sample of tests (e.g., a convenience sample)

5. Study includes a census of tests (e.g., all tests within a given time period), but the time and/or number of tests is insufficient to be representative

6. Study selected a unique and non-representative sample of tests

Method for selecting patients: From the drop down list provided, select the option that best describes which tests or samples were included in the study.

1. Study includes a census of patients (e.g., all patients within a given time period), and the time and/or number of patients is sufficient to be representative.

2. Study includes a randomized or stratified random sample of patients

3. Study used a quasi-random procedure to identify the sample of patients

4. Study used a non-random procedure to identify the sample of patients

5. Study includes a census of patients (e.g., all patients within a given time period), but the time and/or number of patients is insufficient to be representative

6. Study selected a unique and non-representative sample of patients

Describe non-representativeness: Describe, if necessary, sources of deviation from test or sample representativeness. This text field allows the abstractor to record any sample information they feel restricts our confidence that the study results generalize to all tests and populations.

Study Design: Select from the pull-down list the option that best captures the study design on which the findings are based. Note that all comparison designs are included in options 1 – 5, while one group designs are captured in options 6 – 11. If the design is not contained in this list of design options, select option 12 – Other.

Two Group Designs

1. Randomized Controlled Study (assignment at individual level)

2. Group-Randomized Controlled Study

3. Non-randomized comparison study (e.g., Natural Experiments)

4. Case Control (Groups defined by outcome, AKA Retrospective cohort study)

5. Cohort Study (Groups defined by predictor, AKA Prospective cohort study)

One Group Designs

6. Cross-Sectional (e.g., outcome only)

7. Time-series (i.e., One-group over time, multiple measures (e.g., monthly))

8. Before-after

9. Observational study (e.g., Quality Improvement)

10. Individual case study (the "what went wrong" write-ups)

11. Descriptive analysis (e.g., implementation study, feasibility study)

12. Other specify

Additional Information about Study Design / time period / population rating: If option 12 Other is selected above, please include a description of the design this field. If uncertain of the study design, include a description of the design here.

Rate Study Design / time period / population: From the drop-down list provided, rate the uniqueness of the study design, time period, or sample with respect to representativeness. Representativeness is a judgment on whether the results obtained through the study design are likely representative of the results of the practice.

1. Deduct 0 points

2. Deduct 1 point if the sample (either subjects or tests) may not be representative of the likely results of the practice with respect to how the sample was obtained or identified

3. Deduct 2 points if the sample (either subjects or tests) are probably unrepresentative of the results of the practice with respect to how the sample was obtained or identified.

4. Deduct 3 points if the sample is sufficiently unrepresentative based on how it was obtained/identified to clearly nullify the generalizability of the results.

Additional Information about Study Design / time period / population rating: If points were deducted in the previous item, providing a supporting explanation for how the sample is unlikely representative.

Comparator Practice(s) - please describe: Provide a rich description of each comparison practice (copy or paraphrase if necessary). Ideally, each practice could be replicated from description. As such, include in the description any information available about practice: 1) Content and organization, 2) Implementation, 3) Setting (where the practice was implemented), 4) Training, and 5) Required materials and technology. Costs to implement and maintained should be discussed in a later field, under the Practices tab.

Comparator Practice is Ongoing: Please select from the drop-down list of options whether the practice likely continued beyond the period of data collection. (Does not require knowledge of whether it remains in use today.)

Describe Relationship of Comparison Practice to Data Collection: Ideally, data are collected on well established practices and the practice is maintained throughout the entire period of data collection. Please describe the overlap of data collection with practice implementation and practice maintenance. For example, likely statements in this text box could be (1) Practice was well established and maintained throughout data collection, (2) Practice was implemented at the start of data collection and continued throughout the period of data collection, or (3) Practice was discontinued at the end of data collection.

Quality of randomization: The purpose of randomization is to control for artifact due to historical, maturational, selection, etc. In an ideal world, all sources of pretest differences are eliminated, or at least relegated to error variance through the law of large numbers. If too few units are assigned, or if there are groupwise differences (e.g., assignment at the group level) it is likely that not all differences are controlled for and differences documented by a statistical treatment may be due to third factors not controlled by randomization. Attrition, assessed later, may also affect the quality of randomization if the attrition is differential (varies by some substantive source). Thus, the critical issue surrounding the quality of randomization is whether the randomization process effectively controlled for all substantive factors other than the practice that could affect the veracity of the finding.

1. All relevant baseline differences controlled for by randomization

2. Most relevant baseline differences controlled for by randomization

3. Baseline differences not likely controlled for by randomization

4. Not relevant to design

Attrition: Severe, unexplained, or differential attrition after assignment / agreement to participate may invalidate randomization and may adversely affect generalizability from an original study sample. Attrition may also be considered when the expected number of units in an analysis (subjects, samples, or tests) are not included in a study. Select from the drop-down list the option that best describes whether all expected units are included in the analysis:

1. There is minimal attrition represented in this study

2. There is attrition, but authors attest there are no problems resulting from attrition

3. There is attrition, and the authors acknowledge it affected the quality of the randomization

4. There is obvious attrition but it is not discussed or acknowledged by author

Rate Potential Study Bias: From the drop-down list provided, rate the extent to which you believe that the study design, period of measurement, and/or sample selection introduced bias to the results. Bias is systematic error that either reliably enhances *or* supresses a finding. It is different from error, which reduces the quality of measurement, but not in any systematic fashion (can produce findings that are both larger and smaller than the true score).

1. Deduct 0 points

2. Deduct 1 point if the study design, time period and sample selection methods may introduce a study bias that would substantially affect results (i.e., may produce study results interpreted as inconsistent with the true results)

3. Deduct 2 points if the study design, time period and sample selection methods are likely to introduce a study bias that would substantially affect results (i.e., would likely produce study results interpreted as inconsistent with the true results)

4. Deduct 3 points if there is reason to believe that the study characteristics can not produce results representative of the practice

Additional information about Potential Study Bias rating: If points were deducted in the previous item, providing a supporting explanation for how the finding is likely biased.

Total Points Deducted – Study: Using the information summarized by the three previous Study Rating Items (Rows 13, 23, and 32; Items Rate Facility Description / Study Setting, Rate Study Design / time period / population, and Rate Potential Study Bias respectively), record in this field the maximum number of points deducted in any of the previous three Study rating fields. For example, if rows 13 and 23 are each zero, but row 32 equals 3 points deducted, then the total points deducted would be 3. If each of the three previous ratings are 2 then 2 points would be selected for this item.

3 points deducted	2 points deducted	1 point deducted	0 points deducted

Table 3: Practice.

This table provides information on the practices described in the study. These practices are linked through the study identifier to their bibliographic references and via the outcome measures to the results database. Note that a study may summarize multiple practices, contrasting them either directly (e.g., RCTs), or indirectly (e.g., error rates associated with each practice). In this database, code each practice separately using a new column for each practice tested in the study.

Short Name for Study Citation: Topic_Practice_Author_Year: Auto-entered using data recorded in the Biblio field of the same name.

Topic Area – Practice: Auto-entered using data recorded in the Biblio field of the same name.

Practice Name: Assign a short, standardized name for the practice being documented. In the event that a appropriate short name has not been pre-determined, reviewers need to assign one.

Practice - please describe: Provide a rich description of practice (copy or paraphrase if necessary). Ideally, the practice could be replicated from description. As such, include in the description any information available about practice: 1) Content and organization, 2) Implementation, 3) Setting (where the practice was implemented) and population, 4) Training, and 5) Required materials and technology. Costs to implement and maintained and staff responsible and implementing the practice should be discussed in later fields in this worksheet.

Practice Start Date (MM/DD/YYYY): Record the month and year that the comparison practice for this study was implemented. Although the program formats the data as MM/YYYY, data must be entered using the MM/DD/YYYY format. If the day and month are not reported, enter your best estimate of the start of data collection.

Practice End Date (MM/DD/YYYY): Record the month and year that the comparison practice for this study ended. Although the program formats the data as MM/YYYY, data must be entered using the MM/DD/YYYY format. If the day and month are not reported, enter your best estimate of the start of data collection.

Practice is Ongoing: Please select from the drop-down list of options whether the practice likely continued beyond the period of data collection.

Please describe any relevant training required for implementation, including problems encountered: Please provide a rich description of staff training needs associated with implementing this practice.

Amount of training required: Enter the option that best describes the amount of training necessary to successfully implement the practice.

1. Training is not discussed.
2. Training needs are modest.
3. Training needs are moderate.
4. Successfully implementing practice requires extensive training and ongoing supervision to implement successfully.

Provide available information on Staffing - Number and type of individuals involved in implementing and carrying out the practice: In addition to staff routinely carrying out the practice, some practices require additional staff or staff with unique abilities (e.g., information technology staff, hardware support, staff with specialized medical training) to develop, plan and/or implement a practice. Include here information on the staff necessary to implement and carry out the practice.

Supplies, Equipment, Space, and Other Resources: Please list any additional supplies, equipment, space, or other resources necessary to implement and maintain the practice.

Costs: Start-up costs and ongoing costs for sustaining the practice: In this field please record any cost data provided on implementing and maintaining the practice. In particular, note costs associated with information provided in the fields above associated with Training, Staffing, and Supplies, Equipment, Space and Other Resources.

Problems with implementation: Please describe in this field any problems or difficulties noted by the authors in implementing the practice. These can include unanticipated or excessive training needs, problems with technology (hardware and/or software issues), unanticipated or excessive implementation costs, and/or other issues related to institutional support for the practice.

Problems sustaining the practice: Please describe in this field any problems or difficulties noted by the authors in sustaining the practice. These can include unanticipated or excessive staffing needs, problems with technology (hardware and/or software issues), unanticipated or excessive maintenance costs, and/or issues related to institutional support for the practice.

Rate Practice Description: The practice should be well enough described to meaningfully distinguish it from alternative practices and provide a clear understanding of its requirements and characteristics (does not require that the description be exhaustive or support exact replication).

Deduct 0 points

Deduct 1 point if the practice and its basic characteristics are not sufficiently identified.

Deduct 2 points (Score 0) points if the practice and its basic characteristics cannot be clearly identified.

Additional Information about Practice Description Rating: If points were deducted for the practice description, provide a rationale for the deduction.

Rate Adequacy of Practice Description: Ideally, seven components of practice description would be addressed: a) content, b) implementation, c) population / setting, d) training, e) requirements, f) cost, and g) staff responsible and implementing. However, detailed information on all components is not necessary to evaluate the effectiveness and feasibility of implementing a practice, and is typically not provided (e.g., cost and training).

Deduct 0 points

Deduct 1 point if an important aspect/component that is likely to critically affect implementation of the practice is not well described.

Additional Information about Adequacy Rating: If points were deducted on the adequacy of the practice description, provide a rationale for the deduction.

Total Points Deducted – Practice: Using the information summarized by the two previous Practice Rating Items (Rows 16 and 18; Items Rate Practice Description and Rate Adequacy of Practice Description respectively), record in this field the maximum number of points deducted in any of the two previous Practice rating fields. The options for this item are:

2 points deducted

1 point deducted

0 points deducted

Additional information: Use this field to record source for any additional information used or additional information useful for understanding the record being coded.

Any problems with record?: Record any additional information on problems/issues/reservations you had with entering your responses to these practice-level items. Also enter any critical additional information that might be necessary to understanding the practice being documented.

Table 4: Outcome Measures Assessed.

Researchers can estimate the result of any given practice in multiple ways. For example, "time to treat" is one way to estimate the impact of critical values reporting while "reduction in errors" might be another. For some practices, the range of appropriate outcome measures may be quite constrained and consistently reported. For other practices, the range of outcome measures may be limited only by the creativity of the researcher. For the same practice, and within the same study, researchers often report the impact of an intervention on multiple outcome measures. Therefore, a new record column is created for each outcome measure reported by a study. For multiple outcome measures in a single study, place the most important/direct effectiveness measure first (left), and the corresponding study outcome measure ratings in the first column only (i.e., although there may be multiple outcome measures, choose the study outcome measure that has the highest rating for face validity and recording method).

These records will be unique for each study. For example, if five studies each report that the measure used to estimate the impact of bar coding is the "error rate" for patient-specimen identification, five separate records would be created, each uniquely linked to its parent study and practice.

Short Name for Study Citation: Topic_Practice_Author_Year: Copy from data recorded in the Biblio worksheet field of the same name.

Outcome Measure: Short name/ description: Provide short name/description of the measure being recorded. Be sure to note salient characteristics AND notable and important variations/deviations from standardized short name outcome definition.

Outcome Measure: Long Description: Please provide a rich description of the outcome measure. This can copied directly from the source material. If in pdf format the material can often be cut and pasted although hard paragraphs will have to be removed prior pasting the content in Excel. To do this, past the language first in MS Word, search and replace paragraphs "^p" with a space " ". Then cut and paste the reformatted text into Excel.

Relevance rating: Each outcome measure requires a relevancy rating. Outcome measures that more directly address the review question are more relevant to its effect size. For an outcome measures to be rated "Less Direct," the reviewer should provide a rationale for how it is likely to impact a direct measure addressing the review question, and make a case for a strong relationship to a direct measure. For "Indirect," the reviewer should provide a rationale for a likely impact on a direct measure with a weak relationship and/or a relationship for an impact on a "Less Direct" measure.

Also reported are measures that should not be interpreted as an outcome of direct interest (i.e., does not directly address the review question objective), but may contribute to judgments of cost or feasibility as they are within the causal field for the practice. Use the drop-down menu to summarize how directly the outcome measure recorded addresses the review question asked. If not-applicable (e.g., a cost or feasibility of implementation item) select "Not Applicable".

1. Direct
2. Less Direct
3. Indirect
4. Not Applicable - e.g., non-effectiveness measures like cost, feasibility, patient and provider satisfaction

Rate Face Validity of Outcome Measure: Face validity asks if the measure likely captures the construct it is intended to measure. Deduct points if the measure as specified/defined and data collected is possibly or likely flawed and/or if the measure does not capture the essence of what it was supposed to measure (e.g., timeliness). Use the following criteria in making your judgment

1. Deduct 0 points/no deduction

2. Deduct 1 point if:

   (1) Measure does not capture well the outcome being estimated OR

   (2) The measure estimates an outcome that is only modestly related to the evidence review question (e.g., provider satisfaction, compared with change in a patient outcome or an error rate)

3. Deduct 2 points if the measure is confounded by:

   (1)The practice itself (that is, the outcome is a direct result of the practice which was not available or applicable to both the comparison practice and the tested practice) OR

   (2) The measure is confounded by the context in which the practice has been implemented (that is, the outcome is unlikely to be clearly attributed to the practice).

Describe Face Validity rating: If points were deducted for the face validity of the outcome measure, please provide a rationale for the deduction.

Recording Methods Used: Researchers can use different methods for recording the results of a test or trial. Some methods may be subject to bias or fraud while others can be quite robust. Indicate here the method used to record data for the measure being coded.

1. Occurrence log
2. Exception reports
3. Incident reports
4. External complain forms
5. Internal quality control instrument
6. External quality control instrument
7. Other recording method - please specify

If "other" recording method, please specify: If “Other recording method” is selected, please supply a description of the recording method used for the measure being coded

Rate Recording Method: The method for recording or documenting practice results should be reliable and accurate. As mentioned above recording methods may be subject to bias or fraud (intentional and otherwise). Reliability refers to the ability of a measure to return the same result each time, given the same underlying phenomena. Accuracy refers to the precision of the measurement tool. Note that a tool can be accurate, but not reliable (e.g., the adage “measure twice, cut once” observes that while the tape is accurate, the observation of the result is not reliable). Use the following scale to rate the quality of the recording method described.

1. Deduct 0 points

2. Deduct 1 point if the recording method is not described OR does not accurately capture all instances of the outcome

3. Deduct 2 points if the method of recording the outcome is unreliable.

Describe Recording Method Rating: If points were deducted for the recording method of the outcome measure, please provide a rationale for the deduction.

Total Points Deducted – Measurement: Using the information summarized by the two previous Measurement Rating Items (Rows 7 and 11; Items Rate Face Validity of Outcome Measure and Rate recording method respectively), record in this field the maximum number of points deducted in any of the two previous Measures rating fields. The options for this item are:

2 points deducted

1 point deducted

0 points deducted

Rank order measures: Using the review question and Relevancy rating (Row 7) as guides, rank order the effectiveness measures so that the measure that most directly addresses the review question is ranked "1", the next most direct measure is rated "2", and so forth. This rating does not assess how responsive the measure is to the review question and is separate from measure quality – it is a within-study judgment assigning a value to each the measures reported relative to other measures reported by the study. For example, if a study reported three "less direct" measures of a review question, the most relevant of these indirect measures would still be ranked "1" and the least direct measure would be ranked "3". The Total Points Deducted for the first ranked item (the measure that is most relevant – "1") will be used to determine the points deducted for measurement quality when developing the overall study quality rating.

Additional information: Use this field to record source for any additional information used or additional information useful for understanding the record being coded.

Any problems with record?: Record any additional information on problems/issues/reservations you had with entering your responses to these practice-level items. Also enter any critical additional information that might be necessary to understanding the practice being documented.

Table 5: Results.

This table provides information about study outcomes of various study arms. The general rule for creating a new record is to create a new record anytime the measure, sample, or practice changes or when recording a different result (e.g., Type of Effect Size changes effectiveness to "cost" or "feasibility"). To be eligible for coding a study must contain an effectiveness finding (basic inclusion criteria). These can be quantitative information (e.g., error rate, time to treat), cost, or qualitative information for feasibility. A study can contain an almost unlimited number of effect sizes (I once coded something in the neighborhood of 1,400 effect sizes for a single epidemiological study of 411 youth), or may contain only a single finding. In any case, a study must have at least one laboratory practice relevant finding to be coded in this review. Therefore, it is strongly recommended that at least one codable finding be identified prior to coding a study.

Each of the columns/records in this section should represent a single practice/outcome measure combination that relates to at least one study population sample group (potentially with sub-groups). The practices, outcome measures and study population groups and sub-groups should all link directly to those identified in the previous sections. No new study populations, practices and outcome measures can be "created" in this section, only the quantitative and feasibility results related to those previously identified and described.

Short Name for Study Citation: Topic_Practice_Author_Year: Copy and paste using data recorded in the Biblio field of the same name.

Topic Area – Practice: Copy and paste using data recorded in the Practice field Topic Area - Practice.

Practice Name: Copy and paste from the Practice section field (Row 5)

Outcome Measure: Copy and paste from the Measure field "Outcome Measure: Short Description."

Sample Breakout: Breakouts are when the author selects a subsample of the data and provides a separate outcome for that subset. An example might be when a study that uses inpatient data provides separate estimates for the ICU and for the emergency room. Another might be when an inpatient study breaks the data on time of day or shift. Note that this is not a

practices comparison, but a comparison of a given practice in different settings, populations, or facilities. To be a coded as a breakout, the break variable should result in mutually exclusive groupings (even if data on only one of the groups is reported).

1. Finding is based on the entire sample

2. Finding is based on a subset of the available data

3. Other

Describe breakout: If findings are based on a subset of the available data, provide a rich description of the subset of data on which the finding is based.

Enter finding page number (if available): Enter page number(s) or other information that would facilitate finding the evidence being summarized in the original document.

Type of finding: Select the type of finding being recorded. All one group trials involving a single point estimate should be entered as a non-comparative (single point only) effect size.

Impact findings

1. Non-comparative Study, Time series (average)

2. Comparison (RCT, GRCT, NRCT, cross-sectional, Case control, PCS, RCS)

3. Pretest-Posttest

4. Descriptive (one group)

Other findings

5. Cost of implementation

6. Feasibility of implementation ("lessons learned")

7. Cost savings/Cost effectiveness

Finding: Because of the diversity of findings and overall training required to accurately code effectiveness estimates, findings including effect sizes will be entered in narrative form at this stage of the review. Always include the number of subjects and tests/samples on which the effect size is based, the estimate of central tendency (e.g., mean, mode, ratio) and any estimate of dispersion provided (e.g., standard deviation, standard error, variance). For all comparison studies. create separate records for each arm of the comparison and a third record for the comparison. For example, a study contrasting error rates for bar coding with error rates for technician education could have an effect size for the comparison, and separate error rates for each of the bar coding and technician education groups. In this example, each of the three estimates should be entered in separate records. For comparisons, be sure to indicate clearly which arm supplies each set of findings. If a multivariate finding is recorded, include all variables in the model.

If the 'effect size' captures information on the cost or implementation of a practice, enter that information in the Finding field. For this information it is worth being efficient. Enter as much feasibility, cost, implementation, or dissemination data in the Finding field as applicable (e.g., one record for each type of non-effectiveness data provided for a study and practice.

Confirm the direction of effect: For all statistical comparisons, confirm the direction of effect. That is, for effectiveness findings indicate whether the comparison favors the treatment group (a positive effect) or the comparison group (a negative effect). Be aware that direction of effect can be mediated by item wording. In other words, a practice can have a positive impact by reducing errors or increasing accuracy. Note that this variable captures the simple direction of effect, not whether the effect is statistically significant. Note also that as a standardized convention, positive impacts will be coded

positively (i.e., when a practice results in a reduction in errors there is a positive effect from the practice). Following this convention, in a pretest-posttest design an increase in error rates over time would be coded as a negative impact of the intervention. Select only one:

1. Not applicable
2. Positive impact from the practice
3. No difference
4. Negative impact from the practice (i.e., comparison group/pretest group returned a more positive result)

Type of statistical test: Enter the type of statistical test summarized in Finding. Select only one:

Quantitative effect sizes

1. N successful
2. Proportion successful
3. Multi-category frequency or %
4. Means and SDs
5. T-value (independent)
6. Probability with N/degrees of freedom
7. Dependent T-test
8. One-way ANOVA (2 groups, 1 degree of freedom)
9. One-way ANOVA (>2 groups, >1 degree of freedom)
10. Factorial Design (MANOVA)
11. Covariance Adjusted (ANCOVA)
12. Chi-square statistic (1 degree of freedom)
13. Chi-square (> 2x2 table)
14. Nonparametric
15. Correlation coefficient (zero-order)
16. Means and variances
17. Means and standard errors
18. Means and t-test (independent)
19. Multi-factor ANOVA
20. Multivariate analysis

Other findings

21. Cost
22. Implementation
23. Feasibility
24. Dissemination

Effect size significance: If a statistical test was conducted, is the result of the test significant? Select only one:

1. Effect size is significant
2. Effect size is not significant
3. Significance not reported

Statistical power test?: Record if a **s**tatistical power test was conducted or statistical power was discussed

1. Power test was conducted and statistical power was 80% or greater
2. Power test was conducted and statistical power was less than 80%

3. Statistical power is mentioned, but results from the power test are not reported

4. Statistical power is not discussed

Rate Sample Sufficiency: Many of the outcomes of interest are rare events. If too few observations are obtained or if the measurement period is insufficient to capture these events the measure may provide an inaccurate representation of the effect of the practice. Even among more common events, there may also be considerable variation in the number or rate of events over time. The period of measurement should be sufficiently long to allow robust estimates of the impact of the practice.

Deduct 0 points

Deduct 1 point if:

(1) Measurement period insufficient to allow robust estimate of practice impact) OR

(2) Statistical power not discussed AND the sample may be too small to allow a robust estimate)

Deduct 2 points if:

(1) The number of subjects and/or measurement period not reported)

(2) The measurement period is likely insufficient for a robust estimate of practice impact) OR

(3) Statistical power is not discussed AND the sample is likely too small to allow a robust estimate of the impact of a practice)

Describe sample sufficiency rating: If points were deducted for the Sample Sufficiency Rating, please provide a rationale for the deduction.

Rate appropriateness of statistical analysis: Determine if the statistical analysis was appropriate for the data.

Deduct 0 points.

Deduct 1 point if:

(1) Data collected during notably different time periods OR

(2) Data do not permit effect size calculation.

Deduct 2 points if:

(1) Results were obtained using different measures or recording practices for two groups

(2) Inappropriate statistical analysis AND insufficient data to allow or verify calculation of an effect size)

Describe appropriateness rating: If the statistical analysis is rated as inappropriate, describe the rationale for the rating in this field.

Rate Uncontrolled Deviations and Results/Conclusions biases: Despite the best efforts of researchers, studies are sometimes contaminated by unforeseen events or circumstances. Record here the extent to which these may apply to the findings being recorded.

Deduct 0 points

Deduct 2 points if:

(1) Results not due to practice; likely related to different practice(s), OR

(2) Unexplained attrition >70% OR

(3) Uncontrolled differential attrition OR

(4) Results/conclusion not supported by or at odds with the work done/presented

Describe bias rating: If the findings were influenced by Uncontrolled Deviations or Results/Conclusions biases, describe the uncontrolled deviations and results/conclusions biases in this field.

Total Points Deducted – Findings: Using the information summarized by the three previous Findings Rating Items (Rows 15, 17, and 19; Items Rate Sample Sufficiency, Rate appropriateness of statistical analysis, and Rate Uncontrolled Deviations and Results/Conclusions biases respectively), record in this field the maximum number of points deducted in any of the three previous Findings rating fields. The options for this item are:

3 points deducted

2 points deducted

1 point deducted

0 points deducted

Effect Size Magnitude Rating: From the list of available options rate whether the finding obtained is substantial, moderate or minimal to none. In contrast to statistical significance, this is a judgment of the clinical significance of the finding. Does the tested practice result in a meaningful change in outcome relative to its comparison?

1. Substantial

2. Moderate

3. None/Minimal

Identify Key Effectiveness Finding: Identify the comparison records that are based on the highest ranked effectiveness measure from Rank Order – Measures (Row 15 of the Measures worksheet). Of the results based on the number "1" rated measure, select the result that provides the most accurate estimate of the impact of the program. All things being equal, this will usually be the result based on the largest sample. However, if the largest sample finding is based on disparate subjects, samples, or subject to other confounds, a finding based on a subset might be more likely representative of the effectiveness of a practice. Use the drop-down list to identify the finding that best represents the effectiveness of the practice.

1. Most representative of effectiveness

2. Not the most representative effectiveness finding

3. Not applicable, not a comparison

4. Not applicable, not an effectiveness finding

The Total Points Deducted for the identified item (the finding that best represents the effectiveness of the practice) will be used to determine the points deducted for findings quality when developing the overall study quality rating.

Justify Key Effectiveness Rating: Please provide a justification for your selection.

Additional information: Use this field to record source for any additional information used or additional information useful for understanding the record being coded.

Any problems with record?: Record any additional information on problems/issues/reservations you had with entering your responses to these effect size-level items.