Family medicine faculty development (FD) has been the focus of textbooks, dedicated journal issues, special conferences, and other scholarly forums in recent years. Program models have been proposed that range from hour-long workshops to comprehensive fellowship programs lasting 2 years or longer. Systematic reviews have highlighted FD's real and anticipated influence in areas such as clinical teaching, research, and organizational leadership and culture. While many aspects of FD appear to be successful, reviews have singled out the limitations of published FD evaluation reports.

Criticism of published FD evaluation research has targeted shortages of published studies, as well as methodological problems with studies that are often limited to measures of participants’ satisfaction and self-assessed learning. Reviews by Irby and Hekelman, Sheets and Henry, Meurer and Morzinski, and Reid et al contend that available program evaluation studies are insufficient to match the interests of stakeholders.

While stronger methods have emerged in recent years, none stand out as a model for outcome assessment.

To address these limits, FD evaluation researchers have been advised to focus on performance outcomes—actual results that benefit individual faculty members, their environments, and their learners. To guide FD evaluators, an outcomes-focused model developed by Kirkpatrick has been recommended and used in pilot studies. Kirkpatrick’s model describes four levels of program outcomes. The first and simplest outcome level is reaction, which focuses on participant satisfaction with the program. The next outcome level is learning, which refers to program influences on learning attitudes, knowledge, and/or behaviors that can be demonstrated in the learning context. Behavior change, the third outcome level in Kirkpatrick’s model, refers to measurably different trainee activities outside of the training environment that can be attributed to FD influence. Finally, the results outcome level focuses on the effect of training on learners’ careers, patients, students, or work environments. As noted earlier, many FD scholars believe that evaluators should focus on the higher evaluation levels of behavior change and results to foster ongoing stakeholder support for faculty development programs.
Consistent with calls for higher-level evaluation studies and Kirkpatrick’s model, we looked for a method to assess FD program outcomes. Bland et al propose the use of program participants’ curricula vitae (CVs) for outcome evaluation. However, no methods specify how to systematically assess CVs for the purpose of program evaluation. The literature in medical FD has also not provided systematic illustrations of CV use in program evaluation.

CVs hold information about many of the activities valued by FD programs and their stakeholders. In our program, for example, all FD enrollees are full-time faculty in our family and community medicine department, and FD program goals are designed to be consistent with the department’s mission statement. Both the FD program and the department recognize the importance of scholarly publication, peer-reviewed presentations at regional and national meetings, and administrative leadership activities. These are the types of products that trainees are guided to enter into their CVs. Therefore, CV assessment is compatible with both FD program and departmental interests.

Using CVs as evaluation records has an additional advantage. Because CV analysis is a noninvasive record-review method, subjects are not required to react to an instrument. Program participants are often inundated with evaluation questionnaires that can be fraught with response bias, frustrating to administer, and time-consuming to complete. Nonreactive evaluation methods such as record (ie, CV) review have the benefit of reducing response biases while eliminating the need for participants to complete another instrument.

Therefore, in this paper we describe a method for using CVs in FD program evaluation. We applied the method in a pilot study of our FD program graduates to illustrate the method’s strengths and weaknesses.

**Methods**

**Subjects and Program Context**

The study subjects were all 17 participants enrolled in the first two cohorts of the Medical College of Wisconsin’s comprehensive 2-year FD program during the years 1991–1993 and 1992–1994. This program was designed to foster the productivity, retention, and promotion of full-time, junior family physician faculty. These junior faculty are recent graduates from family practice residencies, or they enter our faculty after time in private or managed care practice. These individuals are at the rank of instructors or assistant professors and usually have had no formal training for faculty roles.

FD program instruction is designed to build participants’ skills in six domains: research, education, administration, professional communications, technology, and professional academic skills. The overall program length is 2 years. During each month, participants attend a day (about 5 hours) of group workshops and seminars focused on one or more domain areas. Between group meetings, participants meet with mentors and program staff who assist them with career advancement issues and work on academic projects. To successfully complete the program, participants are required to complete an organizational assessment project and implement or submit to a scholarly forum both an educational and a research project report.

A standard procedure of our FD program evaluation is to request updated CVs from faculty members 2 years after completion of their FD program. These post-program CVs were the records used in this study.

To determine the effect of the FD program on participants, we assigned participants’ CV entries to three time frames. The first was the pre-program time frame, which consisted of 2 years (the calendar year prior to participants’ program entry and the calendar year of entry). The second was the mid-program time frame (the full calendar year in the middle of the 2 academic years). The third was the post-program time frame, which consisted of 2 years (the calendar year of participant graduation, as well as the following full calendar year). Therefore, data collection for the first cohort of FD graduates was for CV entries dated during the years 1990–1994; data collected for the second cohort of FD graduates covered the years 1991–1995.

**Instrument Development**

Based on a content analysis model described by Crabtree and Miller, we designed an instrument to code and inventory CV entries. We first developed a template for analyzing CVs, based on the medical school’s recommended CV format. Two program staff jointly analyzed 5 of the 17 CVs using the initial template and revised the template as necessary to accommodate entries from those 5 CVs. The initial template design and pilot results were reviewed with the FD project director to ensure congruence between coding categories and program objectives. This use of a priori and iterative category development followed the method described by Constas. The category development process resulted in two products, the first of which was a list of 10 CV codes that would be applied to the remaining 12 graduates’ CVs (Table 1).

The second product from the category development process was a code book that provided inclusion, exclusion, and clarifying criteria for coding decisions. During the authors’ process of codebook development, a total of eight criteria were established to guide coding decisions that pertained to the 10 codes of interest in this pilot study. These criteria are listed in Table 2.

**Procedures**

After creation of the template and the code book, the remaining 12 CVs were randomly assigned to the two coders, who followed a systematic procedure for coding them. This procedure included identifying with an
“X” any CV entry for which a coding decision was uncertain due to missing or unclear information. Examples of entries initially coded with an X were leadership titles without years of service and entries that appeared to reflect peer-reviewed presentations that did not list secondary authors. Because we were concerned with consistency and accuracy more than independence, coders interacted briefly on four occasions during the coding of these 12 CVs to agree on application of coding criteria listed in Table 2 or to discuss the appropriate use of the X code.

After the coding procedure was followed for all remaining CVs, semi-structured interviews were conducted between the coders and the CV owners. These interviews rely on data review and input by study subjects after initial data collection to enhance the accuracy and thoroughness of qualitative data. Interviews followed a written protocol that consisted of a description of the FD purpose for the study, discussion and resolution of all items coded X, and opportunities for subjects to ask questions about their own CVs. We agreed that member-check interviews would not be used to solicit unreported products. However, if subjects offered new data (e.g., “I just remembered—I did present one additional paper at STFM in Boston that isn’t shown here.”), the product would be entered on their CV and coded if it met inclusion criteria.

Following the interviews, all CV codes (which were entered directly on the participant CVs) were numbered so that each had a unique identifier. Numbers were then transferred to a summary code sheet. Summary sheet rows corresponded with each of the 10 codes; column headers represented the 5 calendar years of interest. Numerical identifiers were then entered in the appropriate “cell” represented by the correct rows and columns.

A final part of the data analysis consisted of a check for inter-rater agreement to determine the degree of concordance between the two CV coders. We assessed inter-rater agreement 8 weeks after all final coding was completed. This delay in inter-rater reliability assessment was due to an unexpected array of CV styles and entries encountered in the first five CVs. We determined that this variability could necessitate codebook changes beyond those made during joint coding of the first five CVs. As noted above, four consultations to clarify

### Table 1

**Codes Developed and Used for Curriculum Vitae Analysis**

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1 and A2</td>
<td>Publication in a peer-reviewed journal as first author (A1) or non-first author (A2)</td>
</tr>
<tr>
<td>B1 and B2</td>
<td>Presentation in a peer-reviewed regional/national meeting as first author (B1) or non-first author (B2)</td>
</tr>
<tr>
<td>E1, E2, and E3</td>
<td>Leadership in regional or national academic medicine organization as directing officer (E1), other officer (E2), or contributing member (E3)</td>
</tr>
<tr>
<td>F1, F2, and F3</td>
<td>Leadership in medical school-wide organization as directing officer (F1), other officer (F2), or contributing member (F3)</td>
</tr>
</tbody>
</table>

Additional codes (e.g., C1, D1) were developed for CV entries such as teaching awards and grant development. For this pilot study, these codes were not used.

### Table 2

**Codebook Criteria to Assist With Coding**

| General criteria | • Use calendar years, e.g., if committee leadership is entered on the CV as 1992–1993, enter for both years.  
|                  | • All entries that are unclear or appear to be incomplete will be coded X (e.g., presentations reported without what appears to be a full list of coauthors are coded X). |
| Criteria for publications | • Peer-reviewed articles do not include book chapters, letters, abstracts, and book reviews. |
| Criteria for presentations | • Presenting via local teleconferences can be coded as peer-reviewed regional/national presentation.  
|                          | • The local Research Forum is not a peer-reviewed regional/national presentation. |
| Criteria for leadership positions | • The leadership position “codirector” is a “directing officer”—E1 if regional or national in scope, or F1 if at the medical school.  
|                                | • The title “coordinator” is coded as an “other officer” (E2 or F2)  
|                                | • Membership in interest groups (e.g., Society of Teachers of Family Medicine Group on Faculty Development) meet coding criteria (E3, F3) only if a contributing member. |

Only the criteria that apply to the 10 codes of interest in this pilot study are shown in Table 2.
coding criteria occurred during independent coding, but these did not result in codebook revisions. To determine the rate of inter-rater agreement we used Cohen’s Kappa, a concordance measure that corrects for chance agreement.\textsuperscript{14}

To check for inter-rater reliability, we randomly selected four unmarked CVs (with no codes or annotations), two from each primary coder. Each of these four CVs was coded independently by the two coders, using the 10 codes of interest for this study and the X code for those items that needed clarification.

Results

\textit{Inter-rater Reliability}

A total of 70 items on all four CVs were coded by one or both of the coders. The code X was used an average of three times for each CV. There was disagreement on 11 codes, meaning that of all CV entries that received a code, 11 were coded differently, or one coded an item while the other did not. The total proportion of agreement was 84%. Using Cohen’s Kappa to correct for the proportion of agreement that was expected by chance, the corrected proportion of agreement is 79%. This proportion of agreement is excellent based on guidelines reported in the literature.\textsuperscript{14}

\textit{Product Totals}

Overall, interviews accounted for correction or initial entry of 39 codes (13% of the total of all 303 product codes). Of these, 34 resulted from X codes that met inclusion criteria, and five came from subjects’ recalling products during the interviews that fit inclusion criteria.

Figure 1 shows that publication rates for combined faculty (n=17) grew from 3 per year before the program, to 5 per year during the mid-program time frame, and 15 per year after the program. Presentations went from 6 products before the program to 29 products during the mid-program year. Presentations decreased in the post-program time frame, averaging 22 presentations annually for the combined 17 CVs. Administrative leadership activities increased from a group total of 24 positions per year before the program to 42 leadership positions in the mid-program time frame and 46 positions annually in the post-program time frame.

\textbf{Discussion}

CV analysis has been proposed as a method for evaluating FD programs.\textsuperscript{4} This paper describes a systematic method for using CVs as records for FD outcome evaluation. The application of this method showed that CV analysis can be an important method for assessing program outcomes, though evaluators should be aware of unexpected findings and limitations.

At least two parts of this evaluation method’s application were unexpected. First, we found greater-than-expected variation in the organization and completeness of the CVs prepared by our junior faculty. CVs were expected to be based on our institution’s standard CV format, and guidance on CV preparation was provided in an FD workshop, but most CVs were organized in nonstandard ways, often distracting from, rather than enhancing, overall quality. We were surprised to find incomplete or unclear entries (about three per CV, coded with X). Also, subjects’ comments and “product recall” during interviews suggested that some activities may have been underreported, especially in the areas of peer-reviewed presentations and committee leadership activities. Our experience suggests that CVs are not fully improved by expert advice, and their shortcomings limit their “as-is” use in evaluation.

The second surprise in this study was our underestimate of the time needed to interview CV owners. The length of the interview was often increased due to faculty member questions about CV organization and

<table>
<thead>
<tr>
<th>Three Categories of Curriculum Vitae Entries From the Pre-, Mid-, and Post-program Time Frames*</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Graph showing publication rates for combined faculty (n=17)" /></td>
</tr>
</tbody>
</table>

* Subjects are two cohorts of faculty development graduates (n=17).

Values are average annual frequencies over 5 calendar years. Because the faculty development occurred over 2 academic years, the years are defined as: years one and two are designated as pre-program years (4 months overlap with program start-up), year three is the mid-program time frame, and years four and five are designated as post-program years (5 months overlap with program conclusion).
whether or not some products should be included. We agreed early in the interview process that faculty consultations on CV development, while outside the scope of the evaluation, were a good use of our time. While we did not use member-check interviews to actively solicit unreported products, such interviews may yield greater numbers of products. However, reliance on interviews for data collection further undercuts the noninvasive benefit of the record-review method.

Other limitations are associated with this evaluation method. CV data are subject to the limitations of self-report data. While guidelines and advice could have standardized data, varying degrees of self-report bias were observed. Second, there are historical threats to this method’s validity. For example, participation in FD activities may have stimulated faculty to revise their strategies for CV entries over the years of interest. Third, we did not include a control group, so we could not compare the productivity of our graduates with other junior faculty, where one would expect to see some productivity increases due to career maturation alone. Further, the productivity of our subjects may not have been related to their participation in our FD program, per se. Rather, it may have reflected their predisposition to become productive faculty members, and their entry into our program was only a marker for that intent. A fourth limit of this method is the lag time between the effort to produce a product and the product’s final entry on the CV. We believe that the 2-year post-program time frame captured many FD-influenced products, but some written projects may have taken longer. Finally, this study did not assess important products included in most CVs, such as teaching, advising, and departmental and community leadership activities. Entries such as these may present additional challenges to the method of CV analysis described here.

Conclusions

Consistent with calls for greater attention to FD program outcomes, this paper described a method for systematically analyzing CVs and illustrated the method by applying it in a pilot study to reveal its strengths and limits.

Based on our experience, we propose two recommendations that can strengthen this method and ease time expenditures necessary for its use. First, participant CVs should adhere to standard format and reporting guidelines. This recommendation has clear implications for faculty developers who should conduct workshops to make participants’ institutional guidelines explicit and provide hands-on, guided practice organizing and formatting CV entries. Second, member-check contact should be brief, possibly done via e-mail, and focused on assessment. While interviews present real development opportunities that should not be ignored, lengthy consultations should be separated from assessment and done at another time.

The outcomes of this pilot study show increases in the products of interest to our stakeholders. We conclude that CV analysis can be a powerful tool for documenting FD outcomes. We continue to apply this method at our own institution, and our experience supports its further application and refinement with other FD programs.

Acknowledgments: This journal article was supported by a faculty development grant, #5 D45-PE 55052, from the Division of Medicine of the Health Resources and Services Administration (HRSA). Its contents are solely the responsibility of the authors and do not necessarily represent the views of HRSA.

This manuscript was presented at the 1997 Society of Teachers of Family Medicine Annual Spring Conference in Boston.

Corresponding Author: Address correspondence to Dr Morzinski, Medical College of Wisconsin, Department of Family and Community Medicine, 8701 Watertown Plank Road, Milwaukee, WI 53226. 414-456-4331. Fax: 414-456-6506. E-mail: jmorzins@mcw.edu.

REcrcences