Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Trainee Digital Growth Chart - An Electronic Health Record Audit Log-Based Dashboard for Assessing Resident Information Gathering Skills: An Observation Cohort Study

Benjamin Rosner

Benjamin.Rosner@ucsf.edu

University of California San Francisco

Jessica Pourian

University of California San Francisco

Robert Thombley

University of California San Francisco

Sascha Cohen

University of California San Francisco

Aaron Tierney

Kaiser Permanente Northern California

Colin Hubbard

University of California San Francisco

Michele Long

University of California San Francisco

Research Article

Keywords: electronic health record, EHR, audit logs, information gathering, assessment, resident, medical education, graduate medical education

Posted Date: March 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8673916/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Information gathering is the foundational skill of clinical reasoning. However, residents and attending physicians lack objective insights into resident competencies in information gathering from the electronic health record (EHR). The EHR audit logs, time stamped records of user activities, may offer a wealth of information about how residents gather patient information.

In this work, we sought to: 1. Establish a benchmark of attending physician expectations of residents' EHR-based information gathering activities at different stages of residency, 2. Develop and pilot a dashboard - the Trainee Digital Growth Chart - to feedback an audit log-derived peer-benchmarked composite measure of information gathering performance to residents and their attending physicians in pediatric hospital medicine (PHM), and 3. Measure if feedback and assessment on objectively derived information gathering performance promotes future improvement.

Methods

We convened a focus group of PHM attending physicians to establish information gathering benchmarks for residents at each stage of their training. We queried EHR audit logs for the information gathering activities of each resident against these benchmarks and piloted the dashboard in a 4-arm observational cohort study during the 2023-24 academic year, with arms based on different degrees of exposure to the dashboard (including none). We used an interrupted time series-like analysis to assess for change in performance before and after Growth Chart exposure, and we conducted a subgroup analysis for interns.

Results

56.9% (29/51) of residents who had historical Growth Chart data available to examine 30 days before or during a PHM block, logged into the Growth Chart, and 24% of these (7/29) also discussed the performance with their attendings (comprising the group with the highest exposure). Interns in this highest exposure group were noted to have significant improvement in future information gathering during daily rounding and inpatient activities.

Conclusion

Through the EHR audit logs, it is possible to passively measure information gathering and to use this information in resident assessment and training. In our study, the most opportune time to positively influence information gathering behavior was during the intern year when EHR information-gathering habits are still being formed.

INTRODUCTION

Central to the development of clinical competence and reasoning is a set of skills described by the U.S. Accreditation Council on Graduate Medical Education (ACGME), the first and foundational of which is information gathering.[1] In modern clinical practice, effectively gathering patient information from the electronic health record (EHR) is a skill that is crucial to develop during training, but one that is challenging to master given that in many institutions, there is no formal curriculum for this in the preclinical years, and once in the clinical phase of training, learners are faced with an extensive volume of structured and unstructured data throughout the EHR. Compounding this challenge is that educators, tasked with teaching and assessing trainees from medical school clerkships through the end of post-graduate training, not only lack objective insights into trainee development of and proficiency in EHR-based information gathering, but may have different expectations of trainees as to what constitutes "effective" EHR-based information gathering at each stage of training. Lacking specific milestones for EHR-related competencies, training to optimally prepare future clinicians for data rich clinical settings may fall short, and the EHR usage behaviors adopted during formative training may then persist throughout one's professional career.[2]

Despite a movement in the U.S. towards assessment of trainees against objective milestones including the Internal Medicine Milestones Project,[3] the ACGME's Pediatric Milestones,[4] and the ACGME's Next Accreditation System,[5] limitations persist. These include, for example, that trainees are infrequently observed by supervisors during patient interactions;[6] poor intra- and inter-rater reliability of standardized assessments;[7] and inference,[6, 8] recall,[9] and implicit gender bias associated with assessment.[10, 11] In response to these challenges, there have been calls for novel assessment approaches that are automated, low burden, objective, realistic, and more reflective of skills required in 21st century medicine.[12–16] While the "big data" contained in the EHR hold great promise for audit and feedback training purposes, little effort has been directed toward leveraging these data to improve graduate medical education.[13]

The EHR audit logs, time-stamped records of user activities (e.g. chart clicks, notes viewed, orders placed, etc.), offer a particularly powerful dataset from which to learn about EHR usage patterns.[17] These logs, or "digital footprints" of actions taken in the EHR have proved to be rich sources of data for research on trainee and clinician behavior,[18] and may offer pragmatic targets of passively collected data to generate objective insights into trainee learning and assessment.[19] However, the modern EHR is a data heavy environment with an extraordinary breadth of structured and unstructured information. The information gathering habits of established clinicians and those still in training are often developed in a piecemeal way, frequently acquired opportunistically during clinical care by observing the practices of peers and near-peers, and without an external reference or benchmarks for what constitutes effective information gathering.[2] This can then lead to suboptimal use, inefficiency, and the need for ongoing support.[2]

In this study, our goals were to: 1. Understand and delineate attending physician expectations of residents' EHR-based information gathering maturation at different stages of residency, 2. Develop and pilot a system, referred to as the Trainee Digital Growth Chart (Growth Chart), using the EHR audit logs to

feed back information gathering activities to residents and their attending physicians, and 3. To understand whether this feedback is helpful in supporting future information gathering.

METHODS

Setting

This study was conducted at the University of California, San Francisco (UCSF), a large, urban, academic medical center with approximately 84 pediatric medicine residents who rotate through the pediatric hospital medicine (PHM) primary inpatient service several times over the course of the 3-year pediatrics residency. This was a quality improvement project and was deemed exempt by the UCSF Institutional Review Board.

Phase 1: Delineating attending physician expectations of resident EHR-based information gathering maturation over residency

Information Gathering Benchmarks

To develop residency-stage specific benchmarks for an audit and feedback system, we convened a focus group of 5 PHM clinical teaching faculty (attending physicians). Attendings were sent REDCap surveys (Supplementary Material 1) asking them to indicate their opinions of the relative importance of different types of patient information to be gathered from the EHR, and whether they would expect residents, at different stages of training, to review this information in each of two contexts: 1. At the time a resident admits a patient (e.g. resident viewing a recent hospital discharge summary, recent clinic notes, vital signs, lab results, emergency department provider notes, radiology reports, etc.), and 2. Throughout daily rounding and inpatient activities (many of the same listed above, additionally including viewing consultant notes, case manager notes, etc.). Attendings were asked to identify the stages of training at which they would minimally expect residents in post-graduate years (PGY) 1, 2, and 3 to personally review these forms of information in the EHR. To avoid the myriad potential number of condition-specific information gathering expectations (e.g. viewing prior microbiology results for a patient being admitted for an infection; viewing prior echocardiography results for a patient being admitted for congenital heart disease), attendings were instructed to set their information gathering expectations for a general, non-disease specific use case of a patient admission, representing the “lowest common denominator” of information gathering expectations. For PGY2 and 3 residents – who often supervise interns – information gathering expectations included those items that the attending physicians expected the supervising resident to independently examine in the EHR, irrespective of whether the information had been shared with the supervising resident verbally by an intern. Attendings were then guided through synchronous focus groups using a modified Delphi approach,[20] to work towards consensus (defined as ≥ 70% agreement) on the minimum set of information they would expect a resident at each stage of training to have reviewed for a given inpatient.

Because there are many ways to view the same data in the EHR, focus group members were additionally asked to identify and achieve consensus on which views within the EHR for each type of data that they considered to be acceptable. For example, lab results pulled passively into progress notes are often only the most common labs, and attendings felt that residents relying on this approach as their only view of lab results would be inadequate. Other views of lab results such as the Summary◊Labs tab within the Epic EHR were considered adequate. In addition, any category of information (e.g. vitals) for which there were one or more EHR modalities for viewing that information that were potentially non-specific for tracking in the audit logs (for example, the Summary◊Daily Rounds view contains an acceptable display of vitals, but also contains many other types of information, rendering the view of this display non-specific for tracking whether trainees viewed the vitals), would need to be eliminated from consideration from information gathering performance measurement. Therefore, any type of data deemed by attendings as “expected” to be viewed by residents was retained for the Growth Chart as long as an unambiguous and traceable means of capturing view events in the audit log for that type of information was available.

After each synchronous focus group, any items not having achieved consensus were disseminated to the focus group members in REDCap surveys for a round of asynchronous consensus building. Synchronous focus groups and surveys alternated until consensus was achieved.

Information Gathering Performance Measurement

To measure resident performance against the consensus benchmarks, Structured Query Language (SQL) queries were written against the EHR audit logs for each half year of residency. Resident information gathering activities were captured from their EHR activities on hospital workstations, desktop, laptop, and mobile instances (e.g. Haiku and Canto). Performance was calculated on a patient-by-patient level for resident admissions, and on a patient-day level for resident daily rounding and inpatient activities (accounting for all activities throughout the day). In both cases, the performance numerator consisted of information gathering actions performed by the resident (as seen in the audit logs) divided by total number of eligible and expected actions to be performed by the resident (given his/her training stage, and the presence of the data). For example, if no Xray was performed, there was no expectation of a resident viewing a report or the image itself. Reciprocally, if there was no expectation set in the benchmarks by the focus group, then even if an Xray was performed, residents would neither be credited for viewing it, nor penalized for not viewing it. Performance across all of a resident’s patients in each of the admission and rounding domains were then summed for each training stage (1st half of PGY1, 2nd half of PGY1, 1st half of PGY2, 2nd half of PGY2, etc.) to yield an aggregate Information Gathering Index (IGI) for each resident at each stage of training.

Phase 2: Pilot Study

Phase 2: Pilot Study

The Trainee Digital Growth Chart was piloted as a quality improvement project during the academic year from July 1, 2023 – June 30, 2024. In the pilot, we sought to assess whether the use of historical aggregate IGI by trainees, attending physicians, or both influences future information gathering performance.

Study Cohort

The study cohort consisted of residents on primary daytime PHM inpatient service, and attendings physicians who rotate with and assess residents on this service. Resident audit log activities in the charts of patients under their care at the time of admission and as part of daily rounds were included.

Pilot Study Design

The pilot study was designed as an observational cohort study rather than a randomized controlled trial because it was felt that the latter might inequitably provide some residents with the benefit of feedback and others without it. The observational cohort study consisted of 4 intervention arms to which residents naturally self-allocated based on their usage behavior. These intervention arms represented incrementally decreasing levels of exposure to the Growth Chart data:

1. Group A: Resident + Attending Exposure – Both the resident and one or more of their attending physicians logged into the Growth Chart to observe historical performance during the target window (defined as within one month of or anytime during the PHM block), and one or both of them attested to discussing the results together,
2. Group B: Resident Exposure Only – Resident logged into the Growth Chart during the target window, but the attending did not (or the attending did log in, but did not discuss with resident),
3. Group C: Attending Exposure Only - Resident did not log into the Growth Chart, but did discuss performance with an attending who did log in during the target window, and
4. Group D: No Exposure - Resident neither logged into the Growth Chart nor discussed their performance with an attending (current practice/control group).

Residents were offered $50 Amazon gift cards as incentives for accomplishing two tasks: 1. Logging in to the Growth Chart at least once in the target window during the study period, and 2. Completing a post-block Qualtrics survey about whether they had discussed their Growth Chart performance with an attending. Attendings were similarly sent surveys about whether they had discussed Growth Chart performance with each resident. Information about the Growth Chart was disseminated through departmental email channels, presentations, attending champions, the Chief Resident, and at each PHM block orientation. Flyers with QR codes for accessing the Growth Chart, for asking questions of study coordinators, and for accessing a UCSF intranet Wiki about the Growth Chart were posted in resident charting rooms.

Pilot Study Outcomes and Analysis

The primary outcome for the study was the change in aggregate IGI between pre- and post-exposure to the Growth Chart.

To assess the impact of the intervention on future information gathering performance, raw IGI scores were normalized onto a 0-100 scale with higher scores representing better performance against the cohort benchmark. We then fit linear mixed effects regression models with random intercepts to account for the clustering of observations by resident. Residual errors were assumed to be autocorrelated with an exponential form. Restricted maximum likelihood was used to estimate variance components and regression parameters.

The model was similar in structure to an interrupted time series analysis (ITSA) with results aggregated at a weekly level. However, unlike a typical ITSA, each resident had their own interruption date because residents rotate through PHM blocks at different times in the year. The interruption date was individualized based on the following:

1. The interruption date for interns (who had no prior data before the Growth Chart was launched at the start of the academic year) occurred after each had completed one full PHM block and generated audit log data before a 2nd PHM block. This would allow them to log in within 30 days before or during a 2nd PHM block for viewing historical performance that might then influence subsequent performance. Therefore, for this cohort, the interruption date was the later of the day they viewed the Growth Chart in the target window (identified through login time stamps) or were exposed to their historical aggregate performance by discussing with an attending within the target window (identified by Qualtrics survey responses); or if neither of the above, then the interruption date was the day before the start of their 2nd PHM block (typically used as a dummy interruption date for those in Group D with no exposure, but for analytical purposes required an interruption date).
2. The interruption date for PGY2 and 3 residents (those with retrospectively measured historical audit log data from prior academic year(s) already populating their Growth Charts on day 1 of the pilot) was similar with the exception that historical data were already available by the time of their 1st (and in many cases only) PHM blocks during the study period.

Fixed effects in the model included time, defined as the number of days after (or before) the intervention date for each resident (representing the slope for the reference group in the pre-intervention period); the intervention group (representing group-specific difference in the intercept relative to the reference group at the time of intervention); whether the observation occurred in the post-intervention period (representing the change in intercept (i.e. step) for the reference group post-intervention); interaction of group and time (representing the group-specific difference in pre-intervention slope relative to the reference group); interaction of post-intervention period and group (representing the difference in group-specific change in intercept (step) post-intervention relative to the reference group); the interaction of post-intervention period and time (representing the change in slope for the reference group); and the

three-way interaction of intervention group, intervention period, and time (representing the difference of group-specific post-intervention slope relative to the reference group). Additional parameters estimated from the model, analogous to an interrupted time series analysis, included comparisons of group-specific step changes post-intervention and changes in slope between pre- and post-intervention periods.

Because it was unknown whether the intervention would have an impact on PGY2 and 3 residents, whose information gathering habits may be more deeply formed and more difficult to modify, we conducted a subgroup analysis of the impact of the intervention on PGY1 residents' information gathering activities which are in a more formative stage.

All statistical significance tests were conducted at an alpha level of 0.05, and corresponding confidence intervals were calculated at a 95% confidence level. Analyses were performed using R v. 4.4.1 (R Core Team, Vienna, Austria) and Stata 18.0 (StataCorp, College Station, TX).

RESULTS

Phase 1: Delineating attending physician expectations of resident EHR-based information gathering maturation over residency

Information Gathering Benchmarks and Measurement

Focus group consensus on information that attendings would expect residents at PGY1, 2, and 3 stages of training to personally examine in the EHR at the time of admission and during daily rounding activities are shown in Table 1. This includes the relative importance of each item as ranked by the focus group participants and the variability in importance (as manifested through the standard deviations) of those items.

Table 1
Information gathering expectations

	Overall Importance (0–9), Average (SD)	PGY1 1st and 2nd semester	PGY2 1st and 2nd semester	PGY3 1st and 2nd semester
Resident view of…	Admission			
Vitals	9.00 (0.00)	-1	-1	-1
Lab results	8.67 (0.71)	1	1	1
Radiology report	8.22 (0.97)	1	1	1
Home medications	8.11 (0.78)	1	0	0
Allergies	8.00 (0.76)	-1	-1	-1
Completed or pending specialist MD consulting note(s) in the ED	7.89 (1.05)	1	1	1
Admission orders, if not the one placing the orders	7.78 (1.48)	1	1	1
Medications administered in ED	7.78 (1.30)	1	0	0
Reconciliation of home medications	7.67 (1.00)	1	0	0
Completed or pending ED physician note(s)	7.56 (1.13)	1	0	0
Any hospital discharge summary within the last 3 years	7.50 (0.92)	0	0	0
Problem list	7.50 (1.29)	-1	-1	-1
Any outpatient encounter note within last 3 Years	7.12 (0.64)	0	0	0
PACS imaging	5.75 (1.83)	0	0	0

Consensus on information gathering expectations by a focus group of 5 PHM attending physicians for residents at each year of PGY1, 2, and 3 ranked by descending importance (9 = most important, 0 = least important). “-1” indicates that an item was not unambiguously measurable in the audit logs and was therefore not included in the Growth Chart. “0” indicates that the item was measurable but that even if important, the attendings would not necessarily penalize a resident for not having viewed it at a given training stage and was therefore not included. “1” indicates that the item was both measurable and expected at a given training stage, and was included in the performance measurement. Abbreviations: SD – Standard Deviation; ED – Emergency Department; PACS - Picture Archiving and Communication System; DNR/DNI – Do not resuscitate/Do not Intubate; POLST – Physician Orders for Life Sustaining Treatment (a form that indicates a patient’s resuscitation wishes).

	Overall Importance (0–9), Average (SD)	PGY1 1st and 2nd semester	PGY2 1st and 2nd semester	PGY3 1st and 2nd semester
Mark home medications as reviewed	5.56 (1.74)	1	0	0
Mark history as reviewed	5.00 (2.06)	0	0	0
Mark allergies as reviewed	4.56 (1.74)	1	0	0
DNR/DNI status	4.75 (2.63)	0	0	0
Other details on POLST		0	0	0
Advanced directive or advanced care planning documentation		0	0	0
History (medical, surgical, family, social)	3.00 (2.52)	0	0	0
Resident view of…in the last 24 hours	Morning rounds and daily activities			
Vitals	8.90 (0.32)	-1	-1	-1
Radiology reports	8.40 (0.70)	0	0	0
Lab results	8.40 (0.97)	1	1	1
MD specialty/consult notes	8.30 (0.82)	1	1	1
Orders	7.50 (1.96)	1	1	1
Active medications	7.10 (1.66)	-1	-1	-1
Medications administered	7.10 (1.66)	0	0	0
Ins/Outs	7.00 (1.41)	-1	-1	-1
Social work notes	6.80 (1.55)	0	1	1
Respiratory therapy notes	6.50 (2.50)	0	0	0

Consensus on information gathering expectations by a focus group of 5 PHM attending physicians for residents at each year of PGY1, 2, and 3 ranked by descending importance (9 = most important, 0 = least important). “-1” indicates that an item was not unambiguously measurable in the audit logs and was therefore not included in the Growth Chart. “0” indicates that the item was measurable but that even if important, the attendings would not necessarily penalize a resident for not having viewed it at a given training stage and was therefore not included. “1” indicates that the item was both measurable and expected at a given training stage, and was included in the performance measurement. Abbreviations: SD – Standard Deviation; ED – Emergency Department; PACS - Picture Archiving and Communication System; DNR/DNI – Do not resuscitate/Do not Intubate; POLST – Physician Orders for Life Sustaining Treatment (a form that indicates a patient’s resuscitation wishes).

	Overall Importance (0–9), Average (SD)	PGY1 1st and 2nd semester	PGY2 1st and 2nd semester	PGY3 1st and 2nd semester
Nutrition notes	6.30 (1.57)	0	0	0
Speech therapy notes	6.10 (2.33)	0	0	0
Home medications at least once during inpatient stay if not trainee who did admission	6.10 (2.51)	1	1	1
Home medications reconciled at least once during encounter if unreconciled	6.10 (2.51)	1	0	0
DNR/DNI status at least once during inpatient stay if not trainee who did admission	6.00 (2.00)	1	0	0
POLST (if present) at least once during inpatient stay if not trainee who did admission	6.00 (2.00)	0	0	0
Physical therapy /occupational therapy notes	6.00 (2.00)	0	0	0
Wound consult notes	6.00 (2.11)	0	0	0
Case manager notes	5.40 (2.32)	0	1	1
Allergies at least once during inpatient stay if not trainee who did admission	5.40 (2.63)	-1	-1	-1
PACS images	5.10 (1.79)	0	1	1
Mark home medications as reviewed if not already done	5.00 (2.79)	0	0	0
View of problem list	4.56 (2.60)	0	0	0

Consensus on information gathering expectations by a focus group of 5 PHM attending physicians for residents at each year of PGY1, 2, and 3 ranked by descending importance (9 = most important, 0 = least important). “-1” indicates that an item was not unambiguously measurable in the audit logs and was therefore not included in the Growth Chart. “0” indicates that the item was measurable but that even if important, the attendings would not necessarily penalize a resident for not having viewed it at a given training stage and was therefore not included. “1” indicates that the item was both measurable and expected at a given training stage, and was included in the performance measurement. Abbreviations: SD – Standard Deviation; ED – Emergency Department; PACS - Picture Archiving and Communication System; DNR/DNI – Do not resuscitate/Do not Intubate; POLST – Physician Orders for Life Sustaining Treatment (a form that indicates a patient’s resuscitation wishes).

	Overall Importance (0–9), Average (SD)	PGY1 1st and 2nd semester	PGY2 1st and 2nd semester	PGY3 1st and 2nd semester
Mark allergies as reviewed if not already done	4.10 (2.60)	0	0	0
Mark history as reviewed if not already done	3.60 (1.96)	0	0	0

Consensus on information gathering expectations by a focus group of 5 PHM attending physicians for residents at each year of PGY1, 2, and 3 ranked by descending importance (9 = most important, 0 = least important). “-1” indicates that an item was not unambiguously measurable in the audit logs and was therefore not included in the Growth Chart. “0” indicates that the item was measurable but that even if important, the attendings would not necessarily penalize a resident for not having viewed it at a given training stage and was therefore not included. “1” indicates that the item was both measurable and expected at a given training stage, and was included in the performance measurement. Abbreviations: SD – Standard Deviation; ED – Emergency Department; PACS - Picture Archiving and Communication System; DNR/DNI – Do not resuscitate/Do not Intubate; POLST – Physician Orders for Life Sustaining Treatment (a form that indicates a patient’s resuscitation wishes).

(Table 1 to go here.)

(Currently placed at end of document, due to size, according to author instructions.)

The Growth Chart dashboards on desktop and mobile views are shown in Fig. 1. On the y-axis, the 100% performance mark represents the composite that would be attained by summing all potential focus-group derived benchmark values for a given training stage. 25th -75th percentile bands are also shown for relative performance within a training cohort. Users could also see performance on gathering specific types of information in drill down tables.

Phase 2: Pilot Study

Participation Rates

63 residents (29 PGY1, 7 PGY2, and 27 PGY3) were eligible to participate (i.e. had day-service rotations in PHM blocks during the study period), totaling 91 resident-blocks. Of the 91 Qualtrics surveys emailed to residents associated with these blocks, 80 were opened (87.9% open rate), and 58 were submitted with either partial or full responses (63.7% submission rate). 31 attending faculty rotated on PHM across 123 attending PHM blocks (attending blocks are typically shorter than resident blocks, so there are more of them) and received Qualtrics surveys at the end of each block. 100 of these attending surveys were opened (81.3% open rate), and 57 were submitted with partial or full responses (46.3% submission rate).

Of the 63 residents’ 91 blocks, 51 had historical data in the Growth Chart available to be viewed (All intern 1st blocks had no data to be viewed because interns rotating through PHM for the first time had no historical PHM blocks). Of these 51 blocks, 29 (56.9%) had residents log into the Growth Chart to

examine their own data within target windows. 26 of the 31 attendings (83.9%) accessed the Growth Chart a total of 119 times, examining a total of 197 views, across 47 of 123 attending blocks (38.2%).

Impact of intervention on aggregate IGI

Through their Growth Chart access behavior and discussions about their Growth Chart performance with their attending physicians, residents naturally fell into one of 4 exposure groups as shown in Table 2. The results of the impact of the different exposure levels to the Growth Chart are shown in Fig. 2 (admissions) and Fig. 3 (morning rounds and daily ac tivities). In each of the exposure groups in aggregate across all residents, there are no significant changes with respect to step or slope before and after exposure. (Table 1 of Supplementary Material 2)

Table 2
Intervention arms

	Exposed to historical performance data by logging into Growth Chart during target window*	Exposed to historical performance data by discussing with attending physician during PHM block	Number of residents
Group A	Yes	Yes	7
Group B	Yes	No	22
Group C	No	Yes	3
Group D	No	No	27
Group E	No	Unknown	4

* Target window is 1 month prior to or within the current PHM block. Abbreviations: PHM = Pediatric hospital medicine.

PGY 1 subgroup analysis of impact of intervention on aggregate IGI

For rounding information gathering activities, PGY1 residents in Group A (highest level of exposure) were noted to have significant improvement in post-intervention slope relative to their own pre-intervention slope and relative to the unexposed control Group D. This was not the case for other intervention groups, or for any groups in admitting activities. (Figs. 4, 5; Table 2B of Supplementary Material 2)

DISCUSSION

In this study, we sought to understand whether the data passively collected in the EHR audit logs could be leveraged to gain insight into EHR-based information gathering activities of residents, and whether

such data when supplied in a resident dashboard as part of formative assessment, could influence future information gathering performance. In so doing, we also sought to characterize attending physician expectations of the types of information to be gathered in the EHR by residents at different stages of training.

Phase 1: Information gathering benchmarks and performance measurement

Leveraging the EHR audit logs to understand trainee acquisition of EHR skills over time is not new. For example, one Internal Medicine study found a decrease in EHR time per encounter over the course of the intern year.[21] A separate study observed that information gathering, as a relative percentage of activities increased over training.[22] While most healthcare systems instruct both new trainees and employees in the use of the EHR, this sort of orientation is often technical, brief, and not specific to a learner's specialty.[2] More dedicated EHR learning curricula have been implemented in some training programs,[23, 24] but the concept of what constitutes effective information gathering on admission or during daily inpatient activities, based on an agreed upon set of expectations set by attending physicians, has not been described. With respect to setting appropriate expectations for residents and teaching effective EHR use for gathering patient information based on these expectations, critical gaps remain.[25]

We are not aware of training programs in which a common set of information gathering expectations or "best practices" have been defined or communicated with residents. This may in part relate to our focus group observation that attending physicians themselves rarely share this information with each other and may have relatively high variability in their own expectations of residents. Qualitatively, for example, in the focus groups where we worked towards consensus on information gathering benchmarks, we noted marked variation among attendings about their opinions of the relative importance of certain types of information to be gathered by residents. While variability between attending physicians' expectations for residents to view *some* types of information was low (e.g. the standard deviations around the importance of gathering vital signs and lab results were very low), it was high for others (home medications, allergies, Respiratory Therapy and Wound Consult notes, and the Problem List) as seen in Table 1. Furthermore, quantitatively, idiosyncrasies in attending expectations were apparent. For example, as seen in Table 1, viewing a recent hospital discharge summary was deemed by attendings to be moderately important at the time of doing an admission; yet the focus group would not penalize residents at any stage for not viewing one. While attendings felt that viewing radiology reports – but not PACS imaging – was important for residents at all stages at the time of admission, the inverse was the case for doing so on daily rounds. These apparent inconsistencies not only reflect the great breadth of views held by attendings on what constitutes important information gathering, but manifest the ill-defined expectations for residents as they sequentially rotate with different attending physicians. This can be problematic when positive learning behaviors have been shown to be strongly associated with alignment between learner and attending physician expectations.[26]

Phase 2: Pilot study

Participation rates

Qualtrics survey response rates of 63.7% among residents and 46.3% among attending physicians is relatively high, given our experience with other surveys we have conducted in which response rates, despite substantial effort, often hover around 20%.[27] In addition, Growth Chart login rates among residents and attendings of 56.9% and 83.9% respectively indicate that the majority of users interacted with the Growth Chart in some way.

Impact of intervention on aggregate IGI

As seen in learner aggregate IGI performance, the benchmarks set by attendings substantially exceeded learner performance. While resident performance against the absolute benchmarks may yield insight to help attending physicians either instruct residents about expectations or recalibrate their own expectations for future benchmark-setting, relative performance against a resident's cohort offers an alternative indicator of information gathering strengths and weaknesses.

For both admission and rounding activities, exposure to the Growth Chart did not appear to impact future behavior in aggregate across the full set of participants. However, it is notable that in the PGY1 subgroup analysis, those residents in the highest exposure group (Group A) had significant improvement in post-intervention aggregate IGI relative to both their own pre-intervention IGI and to the unexposed control Group D for rounding and daily inpatient activities. This is notable from an intervention timing perspective, as it is possible that by the time of PGY2 and 3, EHR usage habits are well formed, and a single intervention may be insufficient to change such habits. This would suggest that the ideal audit and feedback intervention opportunity might be in the intern year. It furthermore suggests that behavior change is optimized not when either the resident or the attending view the information in isolation, but when they *both* do, and discuss it. This is consistent with prior studies demonstrating the power of verbal feedback on influencing future performance,[28] and it supports the observation that interpersonal relationships form the foundation of the most meaningful feedback.[29] The fact that the same effect was not observed for admission activities may be due to higher volumes of data informing the IGI for rounding activities than for admission activities, but would require further study.

Limitations

There are a number of limitations associated with this study. First, the benchmarks created by the focus group of 5 attending physicians represented the consensus opinions of that single group and are particular to the division of labor and typical team roles expected of PGY1, 2, and 3 residents on PHM blocks at our institution. The "absolute" scale of the IGI based on these benchmarks is therefore somewhat arbitrary and could vary by focus group composition. For this reason, relative performance, as indicated in Growth Chart 25th – 75th percentile bands, may be more relevant for assessing performance than an absolute scale. Another limitation is that this approach to audit and feedback for

resident assessment was tested at only one institution in one specialty. Within this specialty, mostly PGY1 and PGY3 residents rotate through PHM, therefore limiting the amount that could be learned from PGY2 residents. Thus, this study's potential for generalizability is unknown. Next, the choice of a study design allowing resident natural selection into intervention arms was intentional in order to ensure that no learner was at an assigned disadvantage to educational resources. But this approach limited the sample sizes in some groups.

Conclusions

Information gathering is the foundation of clinical reasoning, but assessment of resident information gathering remains highly subjective. In this study, we measured considerable variability between attending physicians about what information they would expect residents at different stages of training to have gathered at the time of an admission and throughout rounding and daily inpatient activities. This unexpected finding lays the foundation for an important question in graduate medical education: To promote training in the combination of verbal and EHR-based information gathering (as opposed to purely verbal information gathering from patients), should standards be identified and communicated to learners about what EHR sources of information constitute effective information gathering? Healthy discussions of this question among GME educators will help guide future directions.

One of these important future directions continues to be the opportunity to pragmatically leverage the audit logs of the EHR to identify and learn from information gathering behaviors. An important advantage of doing so includes the ability to nimbly adapt data and benchmark standards as digital medicine and medical education evolve. Approaches such as the Growth Chart can be scaled to nearly any specialty or allied health professional training program in which learners gather information from the EHR. Nevertheless, implementing dashboards and driving their effective use in a busy clinical educational context remain key challenges, particularly when it may not be reasonable to offer incentives for doing so outside of a quality improvement initiative. The Trainee Digital Growth Chart is one of a number of electronic audit and feedback dashboards described in the literature whose goal is to enhance graduate medical educational, and to answer the call of the ACGME's core residency requirement to provide residents with feedback containing data and benchmarks on their patient populations.[30] While the Trainee Digital Growth Chart operationalized and delivered on some of these promises, including positively influencing information gathering behavior change for PGY1 residents when their performance data were supplemented with verbal feedback from attending physicians, other areas remain opportune for improvement. If we are to effectively use electronic audit and feedback tools for the purpose of enhancing GME, embedding this type of feedback into the clinical workflow, while aspirational, will be crucial in the future.

Abbreviations

ACGME
Accreditation Council on Graduate Medical Education

DNR/DNI
Do not resuscitate/Do not Intubate
ED
Emergency department
EHR
Electronic health record
GME
Graduate medical education
IGI
Information gathering index
ITSA
Interrupted time series analysis
PHM
Pediatric hospital medicine
PGY
Post-graduate year
PACS
Picture Archiving and Communication System
POLST
Physician Orders for Life Sustaining Treatment
SD
Standard deviation
SQL
Structured Query Language
UCSF
University of California San Francisco

Declarations

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study (#20-31397) was reviewed by the Institutional Review Board of the University of California San Francisco and was deemed to be a quality improvement study, not human subjects research. Consent to participate does not apply as the IRB of the University of California San Francisco determined this not to be human subjects research.

CONSENT FOR PUBLICATION

Not applicable

CONFLICTS OF INTEREST:

None declared.

CLINICAL TRIAL NUMBER

Not applicable

FUNDING

This project was supported by an Innovations Funding for Education grant from The Haile T. Debas Academy of Medical Educators at the University of California San Francisco, and by a National Board of Medical Examiners® (NBME®) Innovation Grant. The project and the views expressed in this publication do not necessarily reflect the position or policy of NBME, and NBME support provides no official endorsement.

Author Contribution

BR conceptualized and designed the study. AT and BR conducted focus groups. RT developed SQL queries to populate the Growth Chart. SC and BR designed Growth Chart dashboard. JP and CH performed quantitative and statistical analyses. ML led the resident pilot study. All authors contributed to, read, and approved the final manuscript.

Acknowledgement

We thank the following individuals for participating in focus groups: Dr. Glenn Rosenbluth, Dr. Suni Kaiser, Dr. Karen Sun, and Dr. Manisha Israni. We thank the PHM Chief Resident Dr. Olivia Molineaux for disseminating awareness of the Growth Chart during the pilot, and the following residents in the co-design of the Growth Chart with UCSF professional designer Jonathan Pugh: Dr. Fatima Barragan, Dr. Danielle Perez, Dr. Denise Powell, Dr. Tim Dyster, Dr. Justin Bullock, and Dr. Erica Evans. We appreciate collaboration from Amy Day Rossa and Suria Sadat of the UCSF Office of Graduate Medical Education, and manuscript preparation assistance from Estelle Martin.

Data Availability

The datasets supporting the conclusions of this article are not publicly available due to trainee privacy reasons but are available from the corresponding author upon reasonable request.

References

1. Holmboe ES, Lobst WF, ACGME Assessment Guidebook. 2020. https://www.acgme.org/Portals/0/PDFs/Milestones/Guidebooks/AssessmentGuidebook.pdf. Accessed 13 Sep 2021.
2. Stroup K, Sanders B, Bernstein B, Scherzer L, Pachter LM. A new EHR training curriculum and assessment for pediatric residents. Appl Clin Inf. 2017;8:994–1002.
3. Competence Milestones for Resident Performance. | ABIM.org. https://www.abim.org/program-directors-administrators/milestones.aspx. Accessed 11 Nov 2024.
4. Accreditation Council for Graduate Medical Education PMW. Pediatric Milestones. Accreditation Council for Graduate Medical Education. 2021. https://www.acgme.org/globalassets/pdfs/milestones/pediatricsmilestones.pdf. Accessed 12 Jun 2025.
5. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system–rationale and benefits. N Engl J Med. 2012;366:1051–6.
6. Howley LD, Wilson WG. Direct observation of students during clerkship rotations: a multiyear descriptive study. Acad Med. 2004;79:276–80.
7. Kalet A, Earp JA, Kowlowitz V. How well do faculty evaluate the interviewing skills of medical students? J Gen Intern Med. 1992;7:499–505.
8. Wigton RS. The effects of student personal characteristics on the evaluation of clinical performance. J Med Educ. 1980;55:423–7.
9. Hauer KE, Holmboe ES, Kogan JR. Twelve tips for implementing tools for direct observation of medical trainees' clinical skills during patient encounters. Med Teach. 2011;33:27–33.
10. Dayal A, O'Connor DM, Qadri U, Arora VM. Comparison of male vs female resident milestone evaluations by faculty during emergency medicine residency training. JAMA Intern Med. 2017;177:651–7.
11. Rand VE, Hudes ES, Browner WS, Wachter RM, Avins AL. Effect of evaluator and resident gender on the American Board of Internal Medicine evaluation scores. J Gen Intern Med. 1998;13:670–4.
12. Arora VM. Harnessing the power of big data to improve graduate medical education: big idea or bust? Acad Med. 2018;93:833–4.
13. Triola MM, Hawkins RE, Skochelak SE. The time is now: using graduates' practice data to drive medical education reform. Acad Med. 2018;93:826–8.
14. Chahine S, Kulasegaram KM, Wright S, Monteiro S, Grierson LEM, Barber C, et al. A call to investigate the relationship between education and health outcomes using big data. Acad Med. 2018;93:829–32.
15. Swing SR. The ACGME outcome project: retrospective and prospective. Med Teach. 2007;29:648–54.
16. Epstein RM. Assessment in medical education. N Engl J Med. 2007;356:387–96.

17. Adler-Milstein J, Adelman JS, Tai-Seale M, Patel VL, Dymek C. EHR audit logs: A new goldmine for health services research? J Biomed Inf. 2020;101:103343.
18. Rule A, Chiang MF, Hribar MR. Using electronic health record audit logs to study clinical activity: a systematic review of aims, measures, and methods. J Am Med Inf Assoc. 2020;27:480–90.
19. Levin JC, Hron J. Automated reporting of trainee metrics using electronic clinical systems. J Grad Med Educ. 2017;9:361–5.
20. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000;32:1008–15.
21. Chen L, Guo U, Illipparambil LC, Netherton MD, Sheshadri B, Karu E, et al. Racing against the clock: internal medicine residents' time spent on electronic health records. J Grad Med Educ. 2016;8:39–44.
22. Chi J, Bentley J, Kugler J, Chen JH. How are medical students using the Electronic Health Record (EHR)? An analysis of EHR use on an inpatient medicine rotation. PLoS ONE. 2019;14:e0221300.
23. Pantaleoni JL, Stevens LA, Mailes ES, Goad BA, Longhurst CA. Successful physician training program for large scale EMR implementation. Appl Clin Inf. 2015;6:80–95.
24. Samadbeik M, Fatehi F, Braunstein M, Barry B, Saremian M, Kalhor F, et al. Education and Training on Electronic Medical Records (EMRs) for health care professionals and students: A Scoping Review. Int J Med Inf. 2020;142:104238.
25. Rajaram A, Hickey Z, Patel N, Newbigging J, Wolfrom B. Training medical students and residents in the use of electronic health records: a systematic review of the literature. J Am Med Inf Assoc. 2020;27:175–80.
26. Zante B, Klasen JM. Learner-centered education: ICU residents' expectations of teaching style and supervision level. BMC Med Educ. 2021;21:411.
27. Barker W, Maisel N, Strawley CE, Israelit GK, Adler-Milstein J, Rosner B. A national survey of digital health company experiences with electronic health record application programming interfaces. J Am Med Inf Assoc. 2024;31:866–74.
28. Porte MC, Xeroulis G, Reznick RK, Dubrowski A. Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. Am J Surg. 2007;193:105–10.
29. Ramani S, Könings KD, Ginsburg S, van der Vleuten CPM. Relationships as the backbone of feedback: exploring preceptor and resident perceptions of their behaviors during feedback conversations. Acad Med. 2020;95:1073–81.
30. Accreditation Council for Graduate Medical Education. Accreditation Council for Graduate Medical Education. Common Program Requirements (Residency). https://acgme.org/Portals/0/PFAssets/ProgramRequirements/CPRResidency2021.pdf. Accessed 20 Jul 2021.

Figures

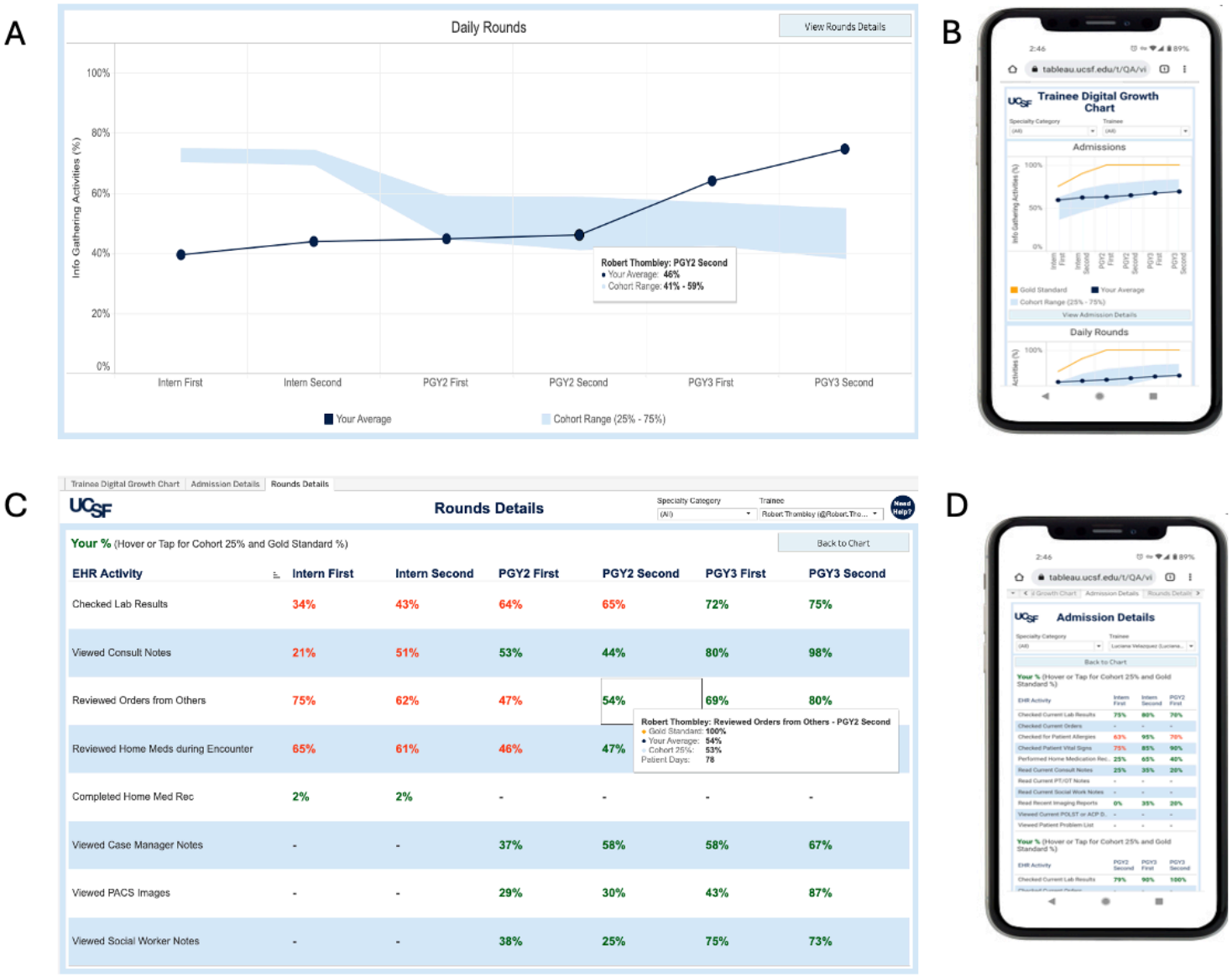


Figure 1

Growth Chart dashboard with sample aggregate IGI over time on A. Desktop, and B. Mobile. Tabular drill down to individual information gathering elements on C. Desktop and D. Mobile. Y-axis shows performance percentage (black dots and lines) against attending physician focus group-derived benchmarks. Black dots and lines show individual average performance over time against 25% - 75% percentile of cohort (blue bands). Red text in drill down tables represents performance below the 25th cohort percentile while green text is 25th percentile or above. Hover-overs provide the resident's average, the cohort range, and the sample size of the number of patient charts from which the data were derived.

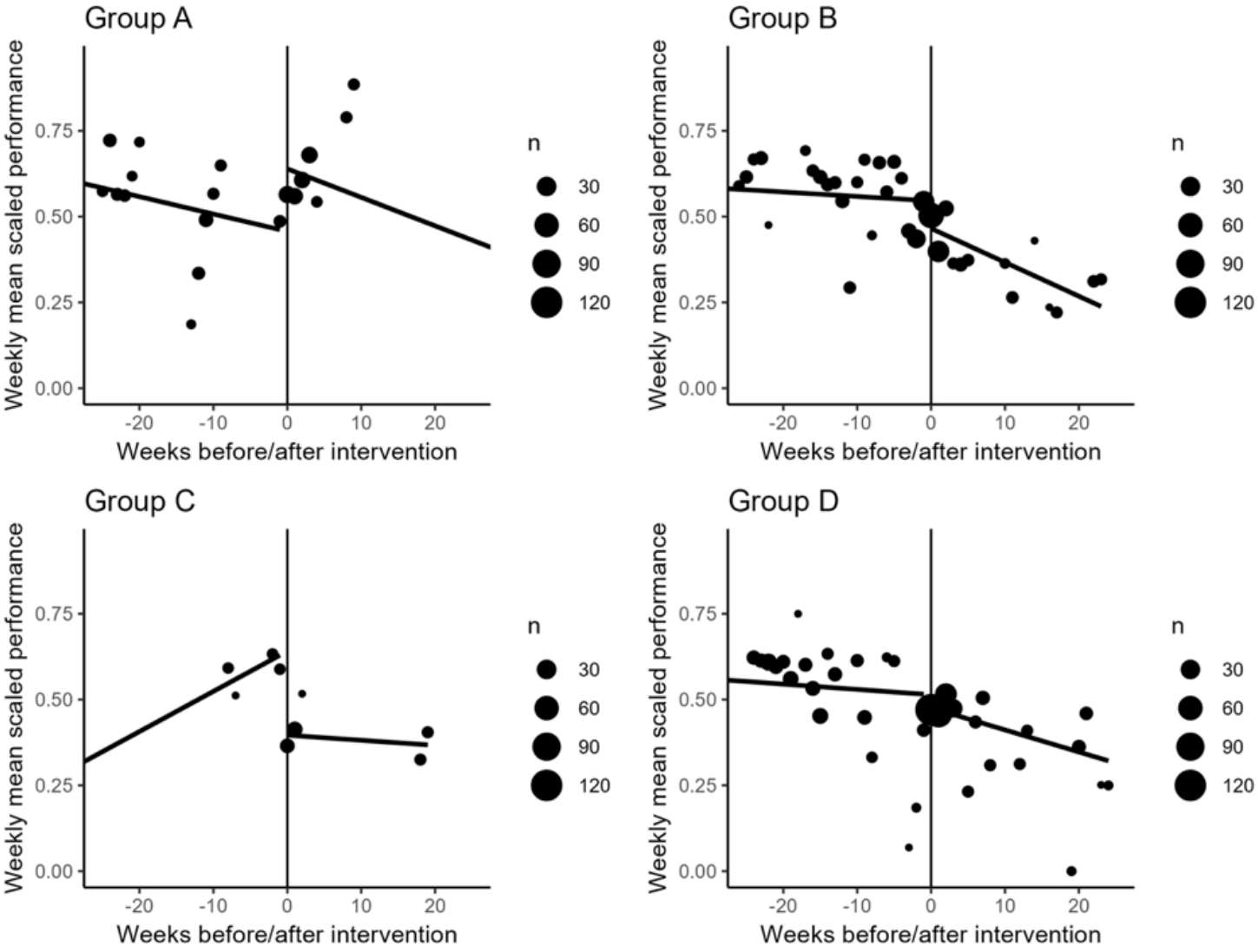


Figure 2

Admission aggregate IGI performance for each of Groups A, B, C, and D (truncated at 25 weeks pre-/post-intervention) across all residents PGY1, 2, and 3. Dots represent mean weekly performance, dot sizes represent number of observations in the given week, lines represent weighted regression line segmented by intervention period. No statistically significant changes. (Table 1A of Supplementary Material 2)

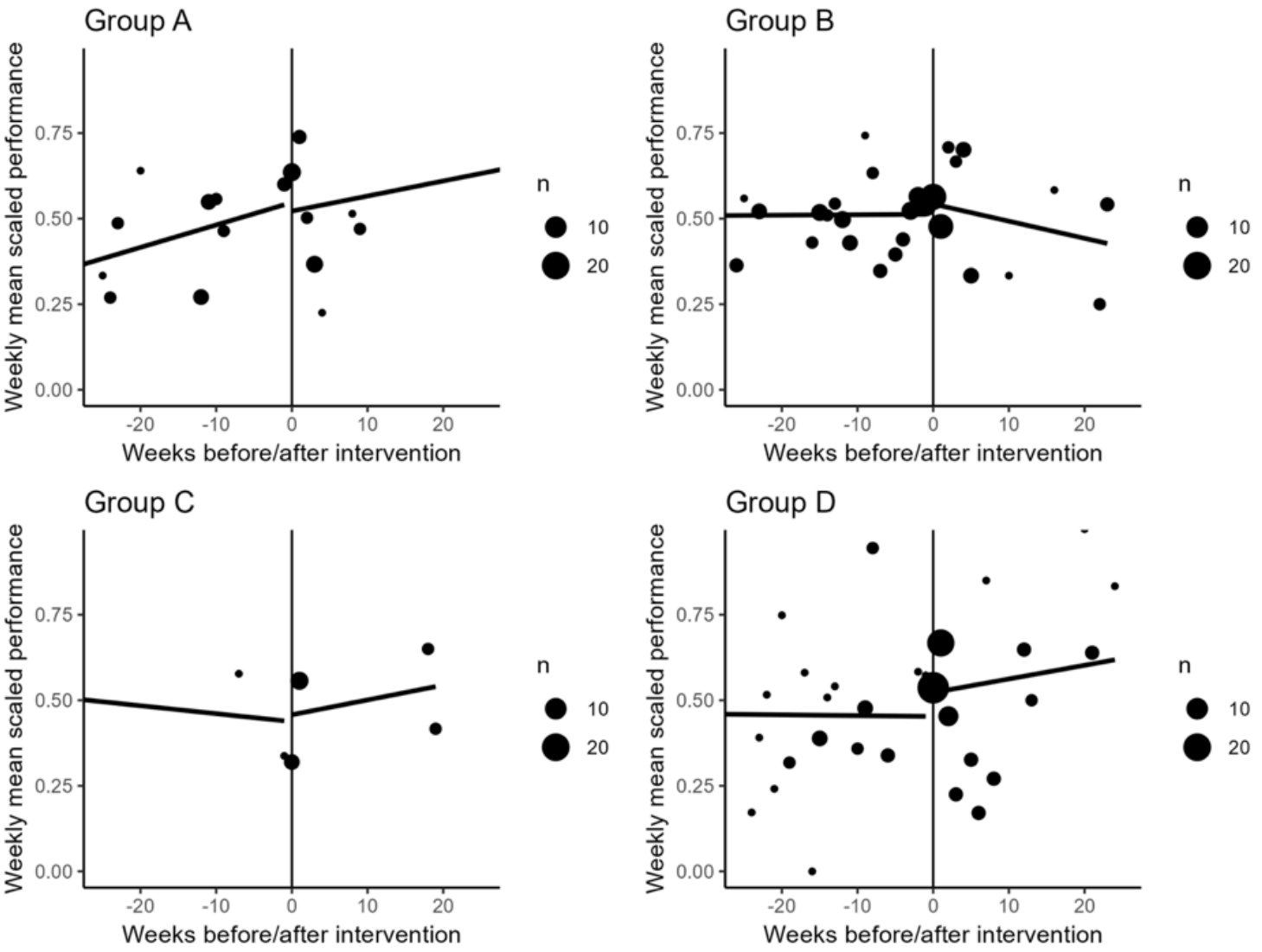


Figure 3

Morning rounding and daily activity aggregate IGI performance over time for each of Groups A, B, C, and D (truncated at 25 weeks pre-/post-intervention) across all residents PGY1, 2, and 3. Dots represent mean weekly performance, dot sizes represent number of observations in given week, lines represent weighted regression line segmented by intervention period. No statistically significant changes. (Table 1B of Supplementary Material 2)

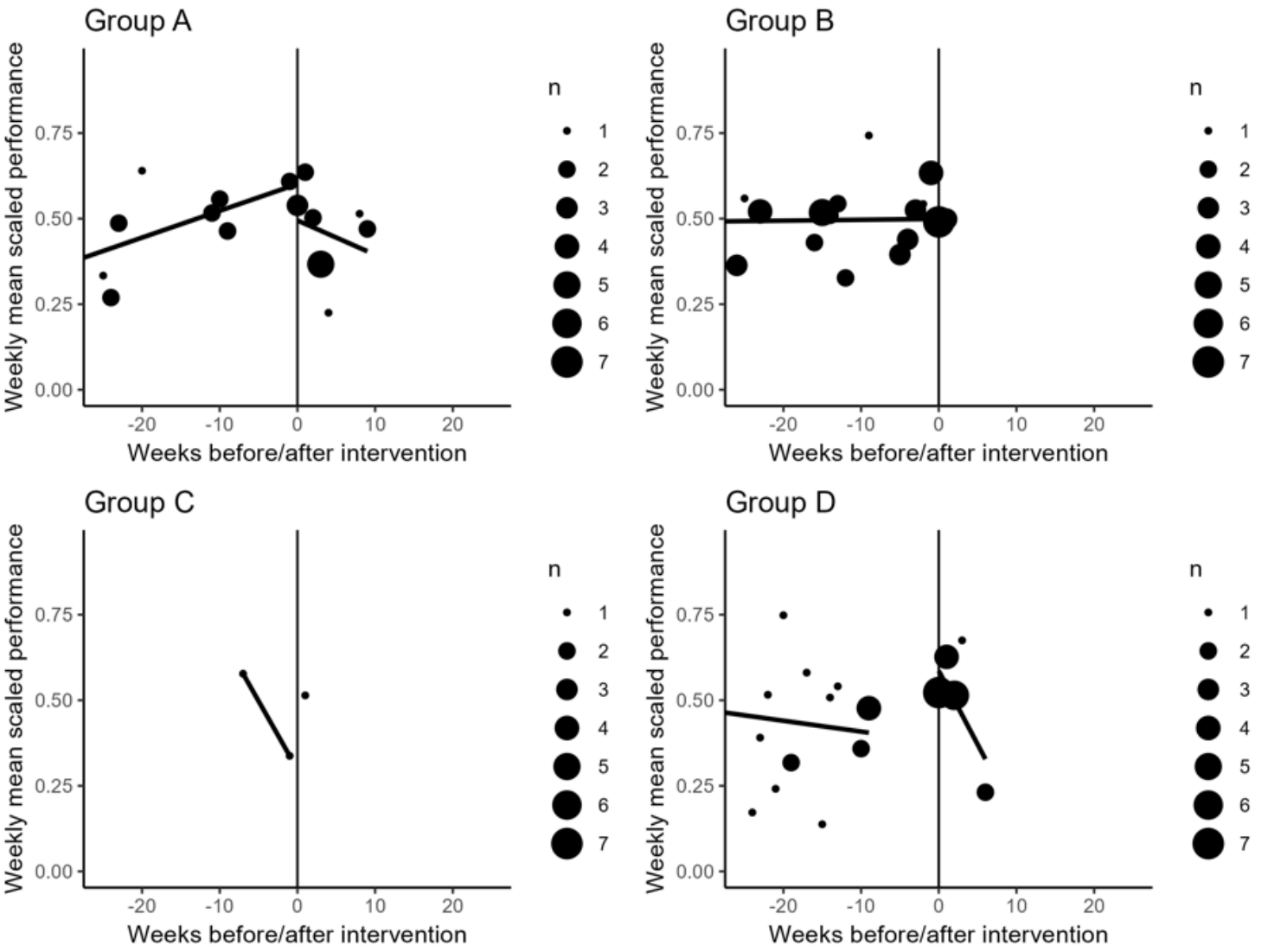


Figure 4

PGY1 subgroup analysis. Admission aggregate IGI performance for each of Groups A, B, C, and D (truncated at 25 weeks pre-/post-intervention). Dots represent mean weekly performance, dot sizes represent number of observations in the given week, lines represent weighted regression line segmented by intervention period. No statistically significant changes. (Table 2A of Supplementary Material 2)

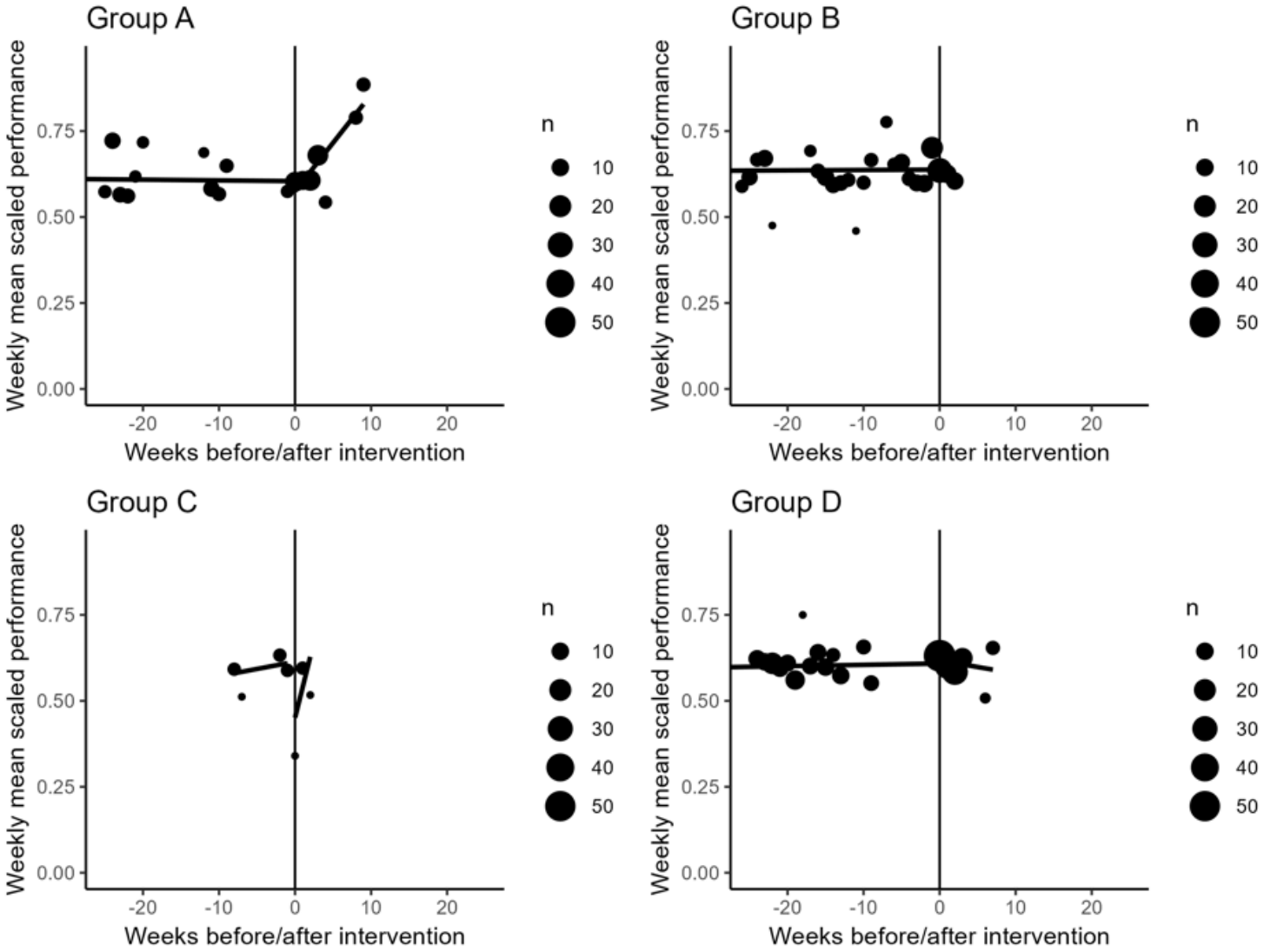


Figure 5

PGY1 subgroup analysis. Morning rounding and daily activity aggregate IGI performance over time for each of Groups A, B, C, and D (truncated at 25 weeks pre-/post-intervention). Dots represent mean weekly performance, dot sizes represent number of observations in given week, lines represent weighted regression line segmented by intervention period. The slope change in Group A pre/post is significantly positive (coefficient 0.0048, *P*<.001), and the post-interruption slope of Group A relative to control Group D is significantly positive (coefficient 0.0037, *P* <.001). (Table 2B of Supplementary Material 2)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RosnerTraineeDigitalGrowthChartBMCSupp11.docx
- RosnerTraineeDigitalGrowthChartBMCSupp2.docx