Using Electronic Health Record Audit Logs to Study Clinical Activity: A Systematic Review of Aims, Measures, and Methods

Adam Rule, PhD¹, Michael F. Chiang, MD¹,², Michelle R. Hribar, PhD¹,²

¹ Department of Medical Informatics and Clinical Epidemiology
Oregon Health & Science University
Portland, Oregon, USA

² Department of Ophthalmology, Casey Eye Institute
Oregon Health & Science University
Portland, Oregon, USA

Corresponding Author:
Adam Rule, PhD
Oregon Health & Science University
Mail Code: BIJC
3181 SW Sam Jackson Park Road
Portland, OR 97239
rulea@ohsu.edu
(206) 291-2533

Keywords: Electronic Health Records; Audit Logs

Word Count: 3,927 / 4,000 words

Note: This is the Author’s Original Version as it was submitted for peer review. The final Accepted Manuscript and Version of Record will differ slightly. This article has been accepted for publication in the Journal of the American Medical Informatics Association Published by Oxford University Press.
ABSTRACT

Objective: To systematically review published literature and identify consistency and variation in the aims, measures, and methods of studies using electronic health record (EHR) audit logs to observe clinical activities.

Materials and Methods: In July 2019, we searched PubMed for articles using EHR audit logs to study clinical activities. We coded and clustered the aims, measures, and methods of each article into recurring categories. We likewise extracted and summarized the methods used to validate measures derived from audit logs and limitations discussed of using audit logs for research.

Results: Eighty-six articles met inclusion criteria. Study aims included examining EHR use, care team dynamics, and clinical workflows. Studies employed six key audit log measures: counts of actions captured by audit logs (e.g., problem list viewed), counts of higher-level activities imputed by researchers (e.g., chart review), activity durations, activity sequences, activity clusters, and EHR user networks. Methods used to preprocess audit logs varied, including how authors filtered extraneous actions, mapped actions to higher-level activities, and interpreted repeated actions or gaps in activity. Twenty studies validated results (23%), but only nine (10%) through direct observation, demonstrating varying levels of measure accuracy.

Discussion: While originally designed to aid access control, EHR audit logs have been used to observe diverse clinical activities. However, most studies lack sufficient discussion of measure definition, calculation, and validation to support replication, comparison, and cross-study synthesis.

Conclusion: EHR audit logs have potential to scale observational research but the complexity of audit log measures necessitates greater methodological transparency and validated standards.
BACKGROUND AND SIGNIFICANCE

Recently mandated logging of electronic health record (EHR) access in audit logs provides a promising resource for researchers to observe clinical activities at scale. Informaticians currently use diverse methods to study both clinical activities and use of health information technology (HIT) including surveys, interviews, and time-motion studies.[1–5] Time-motion studies in particular have seen wide adoption as they avoid many of the biases and inconsistencies of self-report through surveys and interviews. However, the most common form of time-motion study – continuous observation by an external observer – is time-consuming, expensive, and difficult to scale in terms of the diversity, duration, and detail of activity that can be recorded.[1–3] Researchers can scale certain aspects of observational studies with sensors such as Bluetooth beacons and video recorders, but this equipment can be difficult to set up and may provide, depending on the sensor, either a limited stream of data or multi-faceted recordings that require extensive ethnographic analysis.[6,7] Despite the many methods at their disposal, informaticians struggle to observe clinical activity and HIT use accurately, efficiently, and at scale.

Starting in 2005, the Security Rule of the Health Insurance Portability and Accountability Act (HIPAA) required all healthcare organizations to “implement hardware, software, and/or procedural mechanisms that record and examine activity in information systems that contain or use electronic protected health information.”[8] As part of the second stage of the Meaningful Use regulations in 2014,[9] the ONC further clarified that certified EHRs would need to maintain audit logs adhering to the ASTM E2147 standard for tracking HIT use.[10] Due to these regulations, virtually all EHRs in the United States now track at least four pieces of information about every episode of patient record access including who accessed which patient record at what
time and the action they performed in that record such as adding, deleting, or copying information (Table 1). Depending on the vendor, EHR audit logs may track additional information about the computer, user, or record involved in each action, track those actions at different levels of granularity, and give them different names.

<table>
<thead>
<tr>
<th>TIME</th>
<th>USER</th>
<th>RECORD</th>
<th>ACTION</th>
<th>COMPUTER</th>
</tr>
</thead>
<tbody>
<tr>
<td>05/12/2019</td>
<td>SMITHJANE</td>
<td>104738297</td>
<td>Edit Note Section</td>
<td>MED2938</td>
</tr>
<tr>
<td>05/12/2019</td>
<td>SMITHJANE</td>
<td>104738297</td>
<td>Pend Note</td>
<td>MED2938</td>
</tr>
<tr>
<td>05/12/2019</td>
<td>SMITHJANE</td>
<td>104738297</td>
<td>Sign Note</td>
<td>MED2938</td>
</tr>
<tr>
<td>05/12/2019</td>
<td>DOEJOHN</td>
<td>105837489</td>
<td>View Problem List</td>
<td>MED1238</td>
</tr>
<tr>
<td>05/12/2019</td>
<td>DOEJOHN</td>
<td>105837489</td>
<td>View Note</td>
<td>MED1238</td>
</tr>
<tr>
<td>05/12/2019</td>
<td>DOEJOHN</td>
<td>105837489</td>
<td>View Note</td>
<td>MED1238</td>
</tr>
<tr>
<td>05/12/2019</td>
<td>SMITHJANE</td>
<td>107483726</td>
<td>View Patient Summary</td>
<td>MED2938</td>
</tr>
<tr>
<td>05/12/2019</td>
<td>SMITHJANE</td>
<td>107483726</td>
<td>View Patient Summary</td>
<td>MED2938</td>
</tr>
</tbody>
</table>

While originally designed to monitor record access, due to widespread use of EHRs in healthcare EHR audit logs present a unique opportunity to study clinical activities at a scale unachievable with direct observation and with less setup than external sensors. However, like other forms of time-motion study, audit log research is subject to challenges and limitations. Since audit logs are not purpose-built to track workflows they may lack vital contextual information and logged actions may be difficult to map to clinical activities such as chart review or patient exams. Nor do all clinical activities involve EHR use. While EHR audit logs have been used to study diverse clinical activities, there has been little synthesis of the aims of this research, or examination of the variation and validity of measures and methods employed. This lack of synthesis hampers efforts to replicate, generalize, and compare research in areas that may benefit from audit log analysis such as EHR usability and provider burnout.[11–17]

Objective

With this systematic review we identify consistency and variation in the aims, measures, and methods of audit log research. Moreover, we consolidate evidence for the validity of
measures derived from audit logs and limitations of using audit logs to observe clinical activities.

With this review we aim to improve the quality and generalizability of audit log research and provide literature-driven recommendations for the design of future studies to ultimately foster knowledge discovery in areas of critical informatics research.

**MATERIALS AND METHODS**

We identified articles for review by searching PubMed. Since the terms used to describe audit logs vary, we first hand-selected twenty-one audit log articles familiar to us and identified the terms each used to describe audit logs (e.g., access log, usage log, EHR timestamps). Using these synonyms for “audit log” and descriptors of EHRs used in prior systematic reviews,[18,19] we searched PubMed in July 2019 for all literature referencing EHR audit logs (see Appendix for full query). The PubMed query and hand-selection together returned 1775 unique articles, with only one of the hand-selected articles not included in the PubMed results. Through manual title, abstract, and text review, one author (AR) identified 74 of these articles which the met inclusion criteria summarized in Figure 1. Scanning the references of included articles, we identified 12 additional articles which met inclusion criteria, yielding a total of 86 articles for review (Figure 2).

One author (AR) iteratively reviewed and extracted features of each article. These included *study features*: the terms used to describe audit logs, EHR vendor, users studied (e.g., physicians, nurses), duration of study, and reported sample sizes (e.g., number of users, patient records, or encounters studied). This author also extracted each article’s research questions, measures, and data preprocessing methods and together with a second author (MRH) iteratively coded these into a concise set of *aims, measures, and preprocessing methods* used in audit log research. Lastly, one author (AR) extracted and summarized the methods and results of
validation studies and sensitivity analyses reported in reviewed articles as well as limitations discussed of using audit logs for research.

RESULTS

Features of Audit Log Research

The 86 articles included in this review used a variety of terms to describe audit logs in their titles and abstracts (Table 2). Only 31 used terms including the words “audit” or “access” while the remainder referenced more ambiguous EHR data, metadata, timestamps, and logs. Articles also varied in the EHRs, features, and users studied (Table 2). Just over half analyzed audit-logs from commercial EHRs (28 from one vendor, Epic (Verona, WI)). Most articles (66) examined all EHR activity while a minority (20) measured interactions with specific features or data-types such as info buttons, handoff reports, or CT scans. Just over half (47 articles) examined EHR activity in individual departments such as internal medicine, outpatient primary-care, and ophthalmology, while the remainder spanned departments. Only six articles examined EHR use across multiple institutions: four of which were conducted outside the United States and two of which examined interactions with a web-based EHR. Most articles (52) studied all EHR users while the remainder largely studied physician and resident use (31) with only a small number focused on nurses and medical students (3). Most articles (74) reported the length of time studied with the median duration being one year. Articles were less consistent in reporting the number of users, actions, patient records, and encounters studied (Table 3). Just over half of articles were published in 2016 or later (44 articles). See Table 4 for features by article.

Table 2: Features of studies using EHR audit logs to study clinical activity

<table>
<thead>
<tr>
<th>Study Attribute</th>
<th>#</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audit Log Term</td>
<td>31</td>
<td>36</td>
</tr>
<tr>
<td>Audit (e.g., audit log, access log)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generic (e.g., log file, EHR log)</td>
<td>18</td>
<td>21</td>
</tr>
</tbody>
</table>
Table 3: Amount of time, actions, users, patients, and encounters studied varied across articles

<table>
<thead>
<tr>
<th>Time (Months)</th>
<th>Users</th>
<th>Actions</th>
<th>Encounters</th>
<th>Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Studies Reporting</td>
<td>74</td>
<td>50</td>
<td>24</td>
<td>18</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.25</td>
<td>15</td>
<td>20,249</td>
<td>249</td>
</tr>
<tr>
<td>Median</td>
<td>12</td>
<td>155</td>
<td>1,930,620</td>
<td>38,628</td>
</tr>
<tr>
<td>Maximum</td>
<td>120</td>
<td>10,659</td>
<td>118,000,000</td>
<td>3,219,910</td>
</tr>
<tr>
<td>Ref.</td>
<td>PMID</td>
<td>First Author</td>
<td>Year</td>
<td>Audit Log Term</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>------------------</td>
<td>------</td>
<td>----------------</td>
</tr>
<tr>
<td>20</td>
<td>8130443</td>
<td>Michael PA</td>
<td>1993</td>
<td>audit trail</td>
</tr>
<tr>
<td>21</td>
<td>14728157</td>
<td>Cimino JJ</td>
<td>2003</td>
<td>log file</td>
</tr>
<tr>
<td>22</td>
<td>17102263</td>
<td>Chen ES</td>
<td>2006</td>
<td>log file</td>
</tr>
<tr>
<td>23</td>
<td>17238321</td>
<td>Cimino JJ</td>
<td>2006</td>
<td>log file</td>
</tr>
<tr>
<td>24</td>
<td>18693813</td>
<td>Cimino JJ</td>
<td>2007</td>
<td>log file</td>
</tr>
<tr>
<td>25</td>
<td>18308208</td>
<td>McLean TR</td>
<td>2008</td>
<td>metadata</td>
</tr>
<tr>
<td>26</td>
<td>20180439</td>
<td>Bernstein JA</td>
<td>2010</td>
<td>log data</td>
</tr>
<tr>
<td>27</td>
<td>22874273</td>
<td>Ries M</td>
<td>2012</td>
<td>system log</td>
</tr>
<tr>
<td>28</td>
<td>25024755</td>
<td>Hum RS</td>
<td>2014</td>
<td>user log</td>
</tr>
<tr>
<td>29</td>
<td>25954381</td>
<td>Jiang SY</td>
<td>2014</td>
<td>usage log</td>
</tr>
<tr>
<td>30</td>
<td>26958202</td>
<td>Jiang SY</td>
<td>2015</td>
<td>audit log</td>
</tr>
<tr>
<td>31</td>
<td>28808942</td>
<td>Cutrona SL</td>
<td>2017</td>
<td>access/audit log</td>
</tr>
<tr>
<td>32</td>
<td>29696473</td>
<td>Mongan J</td>
<td>2018</td>
<td>audit log</td>
</tr>
<tr>
<td>33</td>
<td>30879186</td>
<td>Epstein RH</td>
<td>2019</td>
<td>access log</td>
</tr>
<tr>
<td>34</td>
<td>-</td>
<td>Asaro P</td>
<td>2001</td>
<td>access log</td>
</tr>
<tr>
<td>35</td>
<td>14728151</td>
<td>Chen ES</td>
<td>2003</td>
<td>log file</td>
</tr>
<tr>
<td>36</td>
<td>15360766</td>
<td>Chen ES</td>
<td>2004</td>
<td>log file</td>
</tr>
<tr>
<td>37</td>
<td>16779018</td>
<td>Clayton PD</td>
<td>2005</td>
<td>audit trail</td>
</tr>
<tr>
<td>38</td>
<td>17213496</td>
<td>Hrpszas G</td>
<td>2007</td>
<td>audit log</td>
</tr>
<tr>
<td>39</td>
<td>18999307</td>
<td>Wilcox A</td>
<td>2008</td>
<td>usage statistics</td>
</tr>
<tr>
<td>40</td>
<td>20442152</td>
<td>Zheng K</td>
<td>2010</td>
<td>audit log</td>
</tr>
<tr>
<td>41</td>
<td>20841655</td>
<td>Bowes WA III</td>
<td>2010</td>
<td>audit log</td>
</tr>
<tr>
<td>42</td>
<td>21292704</td>
<td>Sykes TA</td>
<td>2011</td>
<td>system log</td>
</tr>
<tr>
<td>43</td>
<td>23909863</td>
<td>Park JY</td>
<td>2014</td>
<td>log data</td>
</tr>
<tr>
<td>44</td>
<td>24914013</td>
<td>Ancker JS</td>
<td>2014</td>
<td>EHR data</td>
</tr>
<tr>
<td>45</td>
<td>26618036</td>
<td>Choi W</td>
<td>2015</td>
<td>log file</td>
</tr>
<tr>
<td>46</td>
<td>26831123</td>
<td>Kim S</td>
<td>2016</td>
<td>log file</td>
</tr>
<tr>
<td>47</td>
<td>27332378</td>
<td>Kajimura A</td>
<td>2016</td>
<td>access log</td>
</tr>
<tr>
<td>48</td>
<td>29046269</td>
<td>Kim J</td>
<td>2017</td>
<td>use log</td>
</tr>
<tr>
<td>49</td>
<td>29237579</td>
<td>Lee Y</td>
<td>2017</td>
<td>use log</td>
</tr>
<tr>
<td>50</td>
<td>29295318</td>
<td>Kim J</td>
<td>2017</td>
<td>use patterns</td>
</tr>
<tr>
<td>51</td>
<td>30240357</td>
<td>Shervi EC</td>
<td>2018</td>
<td>access log</td>
</tr>
<tr>
<td>52</td>
<td>30274967</td>
<td>Graham TA</td>
<td>2018</td>
<td>audit log</td>
</tr>
<tr>
<td>53</td>
<td>31183688</td>
<td>Cohen GR</td>
<td>2019</td>
<td>log</td>
</tr>
<tr>
<td>54</td>
<td>24907594</td>
<td>Chi J</td>
<td>2014</td>
<td>audit data</td>
</tr>
<tr>
<td>55</td>
<td>26642261</td>
<td>Ouyang D</td>
<td>2016</td>
<td>electronic audit</td>
</tr>
<tr>
<td>56</td>
<td>26913101</td>
<td>Chen L</td>
<td>2016</td>
<td>audit log</td>
</tr>
<tr>
<td>57</td>
<td>30522828</td>
<td>Cox ML</td>
<td>2018</td>
<td>time data</td>
</tr>
<tr>
<td>58</td>
<td>30815088</td>
<td>Goldstein IH</td>
<td>2018</td>
<td>audit log</td>
</tr>
<tr>
<td>59</td>
<td>30726208</td>
<td>Wang JK</td>
<td>2019</td>
<td>event log</td>
</tr>
<tr>
<td>Ref</td>
<td>PMID</td>
<td>Last Name</td>
<td>First Name</td>
<td>Year</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>-----------</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td>141</td>
<td>30664893</td>
<td>Goldstein IH</td>
<td>2019</td>
<td>audit log</td>
</tr>
<tr>
<td>142</td>
<td>27195306</td>
<td>Senathirajah Y</td>
<td>2016</td>
<td>log file</td>
</tr>
<tr>
<td>143</td>
<td>30137348</td>
<td>Orenstein EW</td>
<td>2019</td>
<td>audit log</td>
</tr>
<tr>
<td>144</td>
<td>22195144</td>
<td>Zhang W</td>
<td>2011</td>
<td>audit log</td>
</tr>
<tr>
<td>145</td>
<td>29481625</td>
<td>Chen Y</td>
<td>2018</td>
<td>interaction patt</td>
</tr>
<tr>
<td>146</td>
<td>21277996</td>
<td>Malin B</td>
<td>2011</td>
<td>access log</td>
</tr>
<tr>
<td>147</td>
<td>22195103</td>
<td>Gray JE</td>
<td>2011</td>
<td>data</td>
</tr>
<tr>
<td>148</td>
<td>24511889</td>
<td>Adler-Milstein J</td>
<td>2013</td>
<td>task log</td>
</tr>
<tr>
<td>149</td>
<td>21292706</td>
<td>Hripcsak G</td>
<td>2011</td>
<td>audit log</td>
</tr>
<tr>
<td>150</td>
<td>29854145</td>
<td>Grando A</td>
<td>2017</td>
<td>event logs</td>
</tr>
<tr>
<td>151</td>
<td>29049512</td>
<td>Read-Brown S</td>
<td>2017</td>
<td>timestamp</td>
</tr>
<tr>
<td>152</td>
<td>28373331</td>
<td>Tai-Seale M</td>
<td>2017</td>
<td>MD Primary</td>
</tr>
<tr>
<td>153</td>
<td>28983817</td>
<td>Arndt BG</td>
<td>2017</td>
<td>event log</td>
</tr>
<tr>
<td>154</td>
<td>30184241</td>
<td>Kannampallil T</td>
<td>2018</td>
<td>log file</td>
</tr>
<tr>
<td>155</td>
<td>22527782</td>
<td>Ben-Assul O</td>
<td>2012</td>
<td>log file</td>
</tr>
<tr>
<td>156</td>
<td>23594488</td>
<td>Ben-Assul O</td>
<td>2013</td>
<td>log file</td>
</tr>
<tr>
<td>157</td>
<td>24692076</td>
<td>Ben-Assul O</td>
<td>2015</td>
<td>log file</td>
</tr>
<tr>
<td>158</td>
<td>26776060</td>
<td>Wanderer JP</td>
<td>2015</td>
<td>audit log</td>
</tr>
<tr>
<td>159</td>
<td>30664473</td>
<td>Soh JY</td>
<td>2019</td>
<td>log</td>
</tr>
<tr>
<td>160</td>
<td>24701327</td>
<td>Gilleland M</td>
<td>2014</td>
<td>usage data</td>
</tr>
<tr>
<td>161</td>
<td>23942926</td>
<td>Hanauer DA</td>
<td>2013</td>
<td>computer log</td>
</tr>
<tr>
<td>162</td>
<td>25074989</td>
<td>Coleman JJ</td>
<td>2015</td>
<td>audit database</td>
</tr>
<tr>
<td>163</td>
<td>30730293</td>
<td>Amroze A</td>
<td>2019</td>
<td>access/audit log</td>
</tr>
<tr>
<td>164</td>
<td>26958173</td>
<td>Chen Y</td>
<td>2015</td>
<td>event log</td>
</tr>
<tr>
<td>165</td>
<td>28269622</td>
<td>Yan C</td>
<td>2016</td>
<td>event sequence</td>
</tr>
<tr>
<td>166</td>
<td>20193841</td>
<td>Shine D</td>
<td>2010</td>
<td>data</td>
</tr>
<tr>
<td>167</td>
<td>27103047</td>
<td>Ouyang D</td>
<td>2016</td>
<td>audit</td>
</tr>
<tr>
<td>168</td>
<td>30625502</td>
<td>Dziomny AC</td>
<td>2019</td>
<td>timestamp</td>
</tr>
<tr>
<td>169</td>
<td>29854253</td>
<td>Wu DTY</td>
<td>2017</td>
<td>audit trail log</td>
</tr>
<tr>
<td>170</td>
<td>29174994</td>
<td>Chen Y</td>
<td>2018</td>
<td>utilization</td>
</tr>
<tr>
<td>171</td>
<td>30807297</td>
<td>Karp EL</td>
<td>2019</td>
<td>event file</td>
</tr>
<tr>
<td>172</td>
<td>26958290</td>
<td>Hribsak M</td>
<td>2015</td>
<td>timestamp</td>
</tr>
<tr>
<td>173</td>
<td>28269861</td>
<td>Hribsak MR</td>
<td>2016</td>
<td>timestamp</td>
</tr>
<tr>
<td>174</td>
<td>29854159</td>
<td>Hribsak MR</td>
<td>2017</td>
<td>EHR data</td>
</tr>
<tr>
<td>175</td>
<td>29036581</td>
<td>Hribsak MR</td>
<td>2018</td>
<td>timestamp</td>
</tr>
<tr>
<td>176</td>
<td>30312629</td>
<td>Hribsak MR</td>
<td>2019</td>
<td>timestamp</td>
</tr>
<tr>
<td>177</td>
<td>27375293</td>
<td>Hirsch AG</td>
<td>2017</td>
<td>audit file</td>
</tr>
<tr>
<td>178</td>
<td>29854142</td>
<td>Goldstein IH</td>
<td>2017</td>
<td>EHR data</td>
</tr>
<tr>
<td>179</td>
<td>29121175</td>
<td>Goldstein IH</td>
<td>2018</td>
<td>timestamp</td>
</tr>
<tr>
<td>180</td>
<td>22574103</td>
<td>Vawdrey DK</td>
<td>2011</td>
<td>audit log</td>
</tr>
<tr>
<td>181</td>
<td>24845147</td>
<td>Chen Y</td>
<td>2014</td>
<td>EHR utilization</td>
</tr>
<tr>
<td>182</td>
<td>25710558</td>
<td>Soulakis ND</td>
<td>2015</td>
<td>record usage</td>
</tr>
<tr>
<td>183</td>
<td>27570217</td>
<td>Chen Y</td>
<td>2017</td>
<td>utilization record</td>
</tr>
<tr>
<td>184</td>
<td>30015537</td>
<td>Yao N</td>
<td>2018</td>
<td>access data</td>
</tr>
<tr>
<td>185</td>
<td>30898243</td>
<td>Durojaiye AB</td>
<td>2019</td>
<td>metadata</td>
</tr>
<tr>
<td>186</td>
<td>31160011</td>
<td>Zhu X</td>
<td>2019</td>
<td>access-log</td>
</tr>
</tbody>
</table>
Aims of Audit Log Research

Most articles used audit logs to study EHR use directly (63 articles, see Table 4 for details by article).[20–82] This included how often providers accessed individual pieces of information,[20–33] patterns of EHR use across features,[34–53,67,74–78], and the total time providers spent using EHRs.[54–60,68–73] More recently, studies began to use audit logs to examine clinical workflows that extend beyond the EHR, using audit log timestamps to mark the boundaries of clinical events (30 articles).[69–98] For example, a few articles calculated resident duty hours using EHR login and logout timestamps, assuming residents log into and out of the EHR near shift boundaries.[79,85–87] Other studies used timestamps to identify the start and end of clinical exams and calculate exam length or patient wait time.[91–98]. Still other workflow studies focused instead on sequences of actions providers took after specific events occurred (e.g. receiving an alert) or their typical workflow when caring for certain patient groups, such as those with complex cardiac conditions.[80–84] A third common use of audit logs was to study care team structure and dynamics (17 articles).[63–69,77,97–105] While a few studies used EHR access to identify care teams for individual patients,[66,99] more used co-access of the same records to identify which providers or departments consistently worked together across patients.[65,100–105]

In addition to these three core aims, many studies collected additional demographic, contextual, or outcome data to model the effect of EHR use on clinical outcomes (12 articles).[42,54,63,64,67,73–76,79,86,104] or the effect of patient, provider, and context on EHR use (15 articles).[31,40,42,46,48,53,57,62,69–71,81,82,96,105] For example, two studies modeled EHR adoption as a function of providers’ demographics and professional
networks.[40,46] Several other studies assessed whether accessing a patient’s historical record decreased their length of stay or chances of being admitted to the hospital.[64,73–76]

Measures of Audit Log Research

Reviewed articles derived a variety of measures derived from audit logs to study these topics including 1) counts of actions captured by audit logs, 2) counts of higher-level activities imputed by researchers, 3) activity durations, 4) activity sequences, 5) activity clusters, and 6) networks of EHR users (summarized in Figure 3, see Table 4 for details by article).

Counts of actions captured directly by audit logs (63 articles) [20–53, 55, 61–67, 74–78, 80, 82–84, 86, 88, 89, 97–105] such as “problem list viewed” were often used to quantify use of specific features such as info buttons and radiology reports. Alternatively, these actions were sometimes aggregated to identify peak periods of EHR use throughout the day or week. Counts of higher-level activities (27 articles) [54–60, 68–73, 79–81, 85–88, 90–96] typically involved first mapping low-level actions to higher-level activities such as chart review and documentation. Alternatively, it might involve looking for significant gaps between actions to identify entire sessions of EHR use or work shifts. These activity boundaries could then be used to compute counts or rates, such as the number of unique EHR sessions across all users in the past month or the percent of encounters where providers reviewed the patient’s historical record. Other studies grouped actions into higher-level activities not to report counts but to compute the duration of those activities including total time devoted to EHR use (30 articles). [24, 31, 54–60, 68–73, 77, 79, 81, 83, 85–87, 90–96, 105]

These first three measures were used to create more complex measures and models of activity, three of which were employed in multiple studies. Eight studies constructed event sequences to identify routine patterns of care and deviations from
them.[35,36,78,83,84,89,96,104] Twelve studies clustered patterns of activity to identify recurring patterns of EHR use, such as which sections of the record providers routinely accessed.[29,30,34–36,48,64,78,83,84,89,104] Finally, eleven articles studying care teams used co-access of a patient’s record to develop networks of users or departments that typically work together.[64–66,68,69,100–105] Across all six measures, there was one significant change in use over time: 47% of articles published since 2016 reported a time duration, whereas only 21% of the articles published before 2016 did so ($\chi^2 = 6.54$, p = 0.01) (Figure 4).

Preprocessing Methods of Audit Log Research

Computing even seemingly simple measures from audit logs such as duration of EHR use is not necessarily straightforward. Yet, less than half of articles (32) discussed how raw audit logs were preprocessed before analysis (see Table 4 for details by article). Fewer still discussed this data wrangling in enough detail to support replication. When reported, common practices included 1) filtering actions, 2) mapping actions to higher-level clinical activities, and 3) selecting criteria to define time-periods. Filtering actions included removing actions that were considered incidental or irrelevant to the study.[30,32,35,40,48,54,56,57,59,75,81,86] For example, one study of medical student EHR use removed short bursts of activity on off-service days, labeling this incidental use.[54] Other studies considered all activity within 24 hours of a patient’s visit as relevant,[40] or only activity in periods with “more than 3 mouse clicks (or 15 keystrokes) or 1700 mouse miles (pixels) per minute”.[56] Another common preprocessing practice was mapping individual actions to higher-level activities such as chart review or documentation.[33,35,36,52,53,64,72,73,82,84,88] While no study reported the actual action-activity mappings, some reported the process used to develop these mappings, which varied as discussed in the next section. A final recurring practice was selecting actions and criteria to
define time-periods.\cite{55,56,58–61,70,71,81,85,94,96,104} This involved defining which actions constituted the start and end of clinical events (such as the first non-login action) and how gaps in activity would be handled. Depending on the research question, meaningful gaps ranged from 5 minutes, which could indicate the user was no longer actively using the EHR,\cite{59} to six hours, which could indicate the end of a shift.\cite{85} Another study identified shifts using a three-step process of 1) identifying distinct shifts based on 4-hour gaps, 2) merging shifts that were less than 7 hours apart which would result in a combined shift length of less than 30 hours, and 3) merging shifts that were less than 2 hours long and would result in a combined shift of less than 20 hours.\cite{87}

Validating Audit Log Measures

Using EHR audit logs to study clinical activity assumes audit logs consistently and accurately track clinical activities and that the methods used to process them into more complex measures are sound. However, a minority of studies reported checking these assumptions through validation or sensitivity analyses. Validation studies, which compared measures derived from audit logs with those obtained through other methods, checked both the mapping of audit log actions to higher-level activities as well as the accuracy of activity patterns or durations derived from audit logs. Of the twenty studies that reported validation analyses, six validated activity mappings and sixteen validated patterns or durations (see Table 4 for details by article).

The six studies that reported validating action-activity mappings used a variety of methods including consensus among two or more researchers,\cite{88} consulting the EHR vendor,\cite{53,56} and direct observation of clinical activities.\cite{72,73,82} Only one of these studies reported the accuracy of mappings, noting that 6.9% of the audit log actions were originally misclassified as representing the wrong activity when compared to direct observation.\cite{72} Of the
sixteen studies that reported validating activity patterns or durations, nine compared them to data self-reported by EHR users, administrators, or the authors. Only seven compared timing data to values obtained through direct observation. Of these, only five reported measure accuracy. Accuracy for EHR time per encounter ranged from overestimating by 43% (4.3 vs 3.0 minutes) to underestimating by 33% (2.4 ± 1.7 vs. 1.6 ± 1.2 min). Measures of appointment lengths which tend to be longer were more accurate, overestimated by just 4% in one study (13.8 ± 8.2 vs.13.3 ± 7.3 min), underestimated by 14% in another (19.4 vs 22.5 min),[71] and overestimated by 29% in a third (24.4 ± 13.0 vs. 18.9 ± 11.0 min).[91] Computing duration data in particular requires making a number of assumptions about what constitutes the start and end of certain activities and how to handle gaps in activity. Four studies reported sensitivity analyses in this vein,[57,59,62,86] such as varying the gap in actions considered idle activity from 5-10 minutes[59] or seeing what impact discarding the first and last 5% of actions in a given work shift had on calculated shift length.[86] None reported a significant change in results due to changing these parameters.

**Challenges and Limitations of Audit Log Research**

Finally, reviewed articles mentioned a few limitations of using audit logs to study clinical activity. First, 20 articles mentioned that audit logs do not provide a full picture of clinical activity as they only capture interactions with the EHR.[20,30,38,51,52,57,64,68,69,72,77,78,85,87,91,92,96,100,104,105] Audit logs do not track phone, pager, or face-to-face interactions nor do they track interaction with paper records and printouts. This may lead to systematic underestimation of interaction or workload. Second, 15 articles noted that gaps between timestamps and multiple concurrent timestamps can be difficult
to interpret.[41,54,58,59,68,70,72,81,82,86,88,90,93–95] For example, does a long gap mean the provider was meaningfully engaged with the EHR that entire time, or had they turned their attention away? Do sequential identical audit log entries (except for the timestamp) represent a repeated or continued activity? Third, seven articles mentioned audit log data were often either too coarse or too detailed for clear interpretation.[20,49,54,69,70,100,103] Logs might capture who accessed a patient record, but not which exact note or result they were viewing. Alternatively, more detailed logs might use different names to track accessing the same piece of information through different screens. It can take researchers substantial time to map these isometric actions to higher-level activities. Lastly, 6 articles noted that audit logs may capture what a user did in the EHR, but data from more qualitative methods such as interviews or observations are needed to understand why.[21,32,38,61,68,81]

DISCUSSION

With this systematic review, we surveyed articles using EHR audit logs to study clinical activities. We found a diverse body of literature employing a range of measures to study EHR use directly, clinical workflows extending beyond the EHR, and care team dynamics. This body of research is growing with over half of reviewed articles published in the last three-and-a-half years. Moreover the increased measurement of duration of EHR use may reflect growing concern over the association between EHR use and provider burnout.[14–17]

Whereas some measures employed in this literature were relatively simple counts of actions tracked explicitly by audit logs, others required researchers to manipulate audit logs in sophisticated ways, generating durations, sequences, clusters, and networks. Many studies glossed over the details of how raw audit logs were preprocessed to compute these measures and even when methods were reported there was significant variation. This variation reflects the
difficulty of interpreting audit logs which requires professional judgement and domain knowledge, such as understanding resident duty hour restrictions.

**Recommendations**

The variability of measures and methods in reviewed articles echoes the variability observed in prior systematic reviews of the time-motion studies in healthcare.[2] It also highlights areas where research using EHR audit logs might improve. We focus our recommendations on four areas: sample size reporting, reporting of methods used to pre-process audit logs, validation and sensitivity analyses, and methodological transparency leading to validated standards.

First, we recommend standard reporting of the time period, number of users, and patient records studied. While most studies report the duration of time studied, not all did. Just over half reported the number of users studied, and far fewer reported the number of patients or encounters analyzed. This use of time to report sample sizes likely reflects the fact that audit log data are routinely queried by time period rather than number of patient records or users desired for analysis. We suggest other reported sample size measures be clinically relevant, such as the number of patient encounters, rather than dataset measures such as number audit log rows, as these are harder to compare across vendors and institutions with different logging practices.

Second, we recommend detailed reporting of steps used to wrangle raw audit log data into measures. Given the variable accuracy of time durations reported in studies that validated them (e.g., from 33% underestimation to 43% overestimation of provider EHR time per appointment), there is still a need to develop more accurate and consistent methods of tracking activities with audit logs. Methods reporting should include any criteria used to filter logs and at least the *process* used to map granular actions into higher-level activities such as documentation.
or chart review. Ideally researchers would also report the exact mapping of actions to activities; however, this may not be feasible given the large number of actions that may map to a single activity or the potential for EHR vendors to consider audit log action names proprietary information. For time durations, we recommend authors report how they handle repeated actions and gaps in activity, as well as how they identify the boundaries of activities, especially if data are missing (e.g., “if a log-out action was missing, we considered the last action before a gap of 2 or more hours the end of the provider’s shift”). We recommend the audit log research community develop standards for reporting more complex measures such as activity sequences, activity clusters, and user networks.

Third, we recommend researchers take more steps to validate and test the sensitivity of their results. Ultimately, the validity of audit log research rests on assumptions that audit logs consistently and accurately track EHR use and clinical activities more broadly. While some methods seem to be approaching parity with direct observation for measuring the duration of longer activities such as patient exams, measures of shorter events such as EHR time per encounter are more varied. Validation may occur in a number of ways including surveys and member-checks, but the gold-standard should remain comparing measures derived from audit logs with those obtained through direct observation. More sensitivity analyses are also warranted as the parameters of methods used to preprocess audit logs may significantly affect results.

Finally, there is a need for greater methodological transparency and validated standards to support replication and synthesis. This includes clear documentation and sharing of data schemas, action-activity mappings, and preprocessing scripts between institutions. We recommend that vendors, institutions, and the audit-log research community work together to share methods and develop validated standards for tracking, querying, and analyzing audit logs.
to compute the diverse measures of clinical activity uncovered in this review. These standards
could in turn support replication and comparison across departments and institutions to identify
consistency and variation in EHR use and clinical workflows between them.

Limitations

This review has a few limitations. First, it does not survey use of all health-information
technology logs, nor even all uses of EHR audit logs. EHR related technologies such as Personal
Health Records, Health Information Exchanges, and mobile health apps often track user activity
with logs similar to EHR audit logs,[106–109] and workflow researchers may use timing data
directly from patient records in their studies (such as admit time or time of placing an order).
EHR audit logs are also routinely used for their primary purpose of access control and several
publications have explored how to use them more effectively for that purpose.[110–114] While
the measures and methods used in these related domains may be similar to those reported in this
review, we scoped our analysis to use of EHR audit logs to study clinical activity to provide
targeting insights for this growing research community. Second, we limited our search to articles
on PubMed which may exclude articles published in computer science or engineering venues not
routinely indexed there. We mitigated this risk by searching the citations of included articles for
relevant references, regardless of venue. Third, our coding process was largely subjective and
performed by a single author. While the authors of each article may not agree with our
classification, we aimed to develop a consistent coding scheme that captured the breadth of the
literature by iteratively defining and applying each category label. Finally, this review likely
reflects a publication bias in which some types of audit log research are more readily published
than others (e.g. workflow studies vs. studies of IT infrastructure needs)

CONCLUSION
EHR audit logs have been used to study a wide range of clinical activities, extending beyond their original purpose of monitoring patient record access. The 86 articles included in this review demonstrate a diverse and growing literature, reflecting researchers’ desire to gather precise data on HIT use and clinical activities at scale. However, the process of turning raw audit logs into insights is complex, requires professional judgement, and varies from study to study, when it is even reported. Moreover, there are relatively few articles in the literature that report testing the validity and sensitivity of audit log measures. This lack of rigor and reporting prevents synthesis and comparison across studies, as well as efforts to improve the accuracy of using audit logs for clinical event measurement. EHR audit logs have untapped potential to support quality improvement and research, but the continued growth of the field will require greater methodological transparency and validated standards to support replication and cross-study knowledge discovery.
COMPETING INTERESTS

The authors have no commercial, proprietary, or financial interest in any of the products or companies described in this article. MFC is an unpaid member of the Scientific Advisory Board for Clarity Medical Systems (Pleasanton, CA), a Consultant for Novartis (Basel, Switzerland), and an initial member of Inteleretina, LLC (Honolulu, HI).

FUNDING

Supported by grants R00LM12238, P30EY10572, and T15LM007088 from the National Institutes of Health (Bethesda, MD), and by unrestricted departmental funding from Research to Prevent Blindness (New York, NY). The funding organizations had no role in the design or conduct of this research.

ACKNOWLEDGEMENTS

Thank you to Julia Adler-Milstein and members of the National Research Network for EHR Audit-Logs and Metadata for help hand-selecting articles to seed this systematic review. Thank you to Julia Adler-Milstein, Genna Cohen, and Nicole Weiskopf for feedback on early versions of this article.

APPENDIX

PubMed Search Query

(`("audit"[All Fields] AND (log[All Fields] OR file[All Fields] OR data[All Fields])) OR (log[All Fields] AND (file[All Fields] OR "event"[All Fields] OR "access"[All Fields] OR "system"[All Fields])))`
OR "usage"[All Fields] OR "activity"[All Fields])

OR "timestamp"[All Fields]

OR "interaction patterns"[All Fields]

OR "utilization patterns"[All Fields]

)

AND

(

("electronic"[All Fields] AND "health"[All Fields] AND record[All Fields])

OR ("electronic"[All Fields] AND "medical"[All Fields] AND record[All Fields])

OR ("computerised"[All Fields] OR "computerized"[All Fields]) AND "medical"[All Fields]

AND record[All Fields])

OR ("electronic health records"[MeSH Terms])

OR ("medical records systems, computerized"[MeSH Terms])

)

AND

English[lang]
REFERENCES


14 Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of

15 Shanafelt TD, Hasan O, Dyrbye LN, et al. Changes in Burnout and Satisfaction With Work-

16 Shanafelt TD, Dyrbye LN, Sinsky C, et al. Relationship Between Clerical Burden and
Characteristics of the Electronic Environment With Physician Burnout and Professional

17 Gardner RL, Cooper E, Haskell J, et al. Physician stress and burnout: the impact of health

18 Hogan WR, Wagner MM. Accuracy of data in computer-based patient records. J Am Med

19 Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality

20 Michael PA. Physician-directed software design: the role of utilization statistics and user
input in enhancing HELP results review capabilities. Proc Annu Symp Comput Appl Med
Care 1993;107–11.

21 Cimino JJ, Li J, Graham M, et al. Use of online resources while using a clinical information
system. AMIA Annu Symp Proc 2003;175–9.


23 Cimino JJ. Use, usability, usefulness, and impact of an infobutton manager. AMIA Annu
Symp Proc 2006;151–5.


25 McLean TR, Burton L, Haller CC, et al. Electronic medical record metadata: uses and


usability evaluation of a new cancer data visualization tool. Stud Health Technol Inform
2012;180:656–60.


Shenvi EC, Feupe SF, Yang H, *et al.* “Closing the loop”: a mixed-methods study about resident learning from outcome feedback after patient handoffs. *Diagnosis (Berl)* 2018;5:235–42.


FIGURE LEGENDS

Figure 1: Article inclusion criteria

The article...

1. is peer reviewed
2. reports original research
3. is a full paper rather than an abstract
4. studies use of an EHR rather than related technologies such as Health Information Exchanges, Personal Health Records, and mobile health apps
5. uses EHR audit log data rather than other logs or EHR data such as admission and discharge timestamps from the patient record
6. involves secondary use of EHR audit logs rather than using audit logs for their original purpose of access control
Figure 2: Article review process

- records from hand search (n = 21)
- records from PubMed (n = 1774)
- search terms
- records after duplicates removed (n = 1775)
- records screened for relevance (n = 1775)
- records excluded (n = 1552)
- full-text articles assessed for eligibility (n = 223)
- full-text articles excluded (n = 149)
- records identified through reference search (n = 12)
- studies included in qualitative synthesis (n = 86)
Figure 3: A) Audit logs track actions EHR users perform in patient records. Here we show a simplified example of an audit log for one provider performing actions (e.g., “View Problem List”) in three different patient records. We have already mapped these actions to three higher-level clinical activities (record review, orders, documentation). B) Audit logs can be used to compute a variety of measures including simple measures such as 1) action counts, 2) higher-level activity counts and 3) activity durations. These base measures may be used to create more complex models and measures such as 4) sequences of activities, 5) clusters of similar activity patterns, and 6) networks of providers based on their access of the same patient records.
Figure 4: Audit log publications over time with publications reporting a time duration highlighted.