The debate regarding anatomy laboratory teaching approaches is ongoing and controversial. To date, the literature has yielded only speculative conclusions because of general methodological weaknesses and a lack of summative empirical evidence. Through a meta-analysis, this study compared the effectiveness of instructional laboratory approaches used in anatomy education to objectively and more conclusively synthesize the existing literature. Studies published between January 1965 and December 2015 were searched through five databases. Titles and abstracts of the retrieved records were screened using eligibility criteria to determine their appropriateness for study inclusion. Only numerical data were extracted for analysis. A summary effect size was estimated to determine the effects of laboratory pedagogies on learner performance and perceptions data were compiled to provide additional context. Of the 3,035 records screened, 327 underwent full-text review. Twenty-seven studies, comprising a total of 7,731 participants, were included in the analysis. The meta-analysis detected no effect (standardized mean difference = −0.03; 95% CI = −0.16 to 0.10; $P = 0.62$) on learner performance. Additionally, a moderator analysis detected no effects ($P > 0.16$) for study design, learner population, intervention length, or specimen type. Across studies, student performance on knowledge examinations was equivalent regardless of being exposed to either dissection or another laboratory instructional strategy. This was true of every comparison investigated (i.e., dissection vs. prosection, dissection vs. digital media, dissection vs. models/modeling, and dissection vs. hybrid). In the context of short-term knowledge gains alone,
INTRODUCTION

Historically, donor dissection (or specimen-based dissection) has been considered the gold standard of gross anatomy intensive training. The introduction of modern alternative teaching modalities, however, raises questions regarding the utility and effectiveness of this traditional anatomy education practice. A number of published studies have attempted to address these questions and, in doing so, have regularly praised the role of dissection. These studies, however, are often grounded in students’ and/or educators’ perceptions (Alagna et al., 1982; Kramer and Soley, 2002), rely on anecdotal evidence (Moore, 1998; Aziz et al., 2002; Miller et al., 2002), are fraught with presumed researcher bias (Winkelmann, 2007), or lack the necessary power to make generalizable conclusions (Winkelmann, 2007). Inevitably, interpreting such studies in isolation, or without the proper methodological context, can result in a skewed representation of the empirical evidence.

There are multiple examples where vague or ambiguous findings regarding the anatomy dissection debate are demonstrated in the literature. One such example is an article by Winkelmann (2007) who posited that traditional dissection showed a slight advantage in improving learning outcomes and anatomical knowledge over prossections. Although this review included 14 studies, it fell short methodologically because a thematic review strategy was used to interpret numerical data (Borenstein et al., 2009). As such, the conclusions drawn about the effects of laboratory-based instructional approaches on student performance were purely subjective and therefore quite limited.

Another example of equivocal findings is a meta-analysis which explored the effectiveness of physical 3D models compared to other anatomy education methods (e.g., 2D images, cadavers, and virtual reality) (Yammine and Violato, 2015). In this analysis, data were derived from only eight studies (totaling 820 learners). Moreover, the study did not disclose whether the meta-analysis had sufficient power for making substantive generalizations. Despite these limitations, Yammine and Violato (2015) purported that the use of physical models, compared to an assortment of educational methods, had a positive effect on overall knowledge outcomes and spatial knowledge acquisition.

A review by Tam et al. (2009) reported on the effectiveness of computer-aided learning (CAL) used within anatomy education. Their analysis of eight quantitative studies and articles of learners’ attitudes concluded that well-designed CAL programs are particularly useful as supplemental learning aids for complex anatomy topics (e.g., dissection areas that are difficult to access or preserve). Their study, however, had insufficient evidence to conclude whether CAL programs are a viable replacement for traditional methods of anatomy instruction.

In the words of Plack (2000), “Medical literature is replete with research and essays debating the pros and cons of dissection versus prossection in the teaching of human gross anatomy.” In contrast to prior studies, the present capstone employs a meta-analytic approach to summarize the efficacy of anatomy laboratory pedagogies (e.g., prossection, digital media, 3D models/modelling, and hybrid approaches) across a collection of studies published over the past 50 years and addresses the methodological limitations found in previous works.

In an era of advanced educational technologies and intensified administrative and/or organizational pressures to streamline resource allocations, the findings of this work may help to substantiate which laboratory-based approach, if any, is more efficacious for anatomy education. Admittedly, as described in a narrative review by Bergman et al. (2011), the inherent nuances and complexities of anatomy education make investigating this topic especially challenging. Our research question asked, “How effective is dissection compared to other anatomy laboratory pedagogies as a medium for teaching gross anatomy?” Based on presumptions surmised from previously published thematic reviews, we hypothesized that traditional dissection using specimens (compared to other laboratory instructional approaches) would demonstrate a slight positive effect on learner performance scores across multiple studies and learner populations. Additionally, we anticipated that the practice of active dissection would be valued/favored over other laboratory modalities by a prevalence of positive perceptions.

MATERIALS AND METHODS

This meta-analysis was conducted in accordance with PRISMA guidelines (Supplement 1 in Supporting Information) for the reporting of systematic reviews and meta-analyses (Moher et al., 2009) and follows recommendations outlined by Cook and West (2012). Published articles (including ahead of print publications), dissertations, and meeting abstracts were searched across a 50 year range between January 1965 and December 2015 on PubMed, CINAHL, PsycINFO, ERIC, and Dissertations Proquest & Theses A&I. The search query (Supplement 2 in Supporting Information) combined the concepts of dissection/prosection/noncadaveric resources/peer-teaching with...
the topics of anatomy and education within the areas of student performance or perceptions. Medical subject headings and key terms related to each of these areas were incorporated in the search. Articles that underwent full-review were also hand-searched to identify relevant studies omitted by the electronic search.

**Eligibility Criteria**

Studies were included for preliminary review if they evaluated the educational effectiveness of traditional dissection compared to either prosecution or nonspecimen-based learning resources in nonexpert learners. Nonexpert learners were defined as cohorts of students (e.g., first year medical students) with collective inexperience in intensive laboratory-based anatomy. Experienced learners (e.g., fourth year medical students or residents) were excluded as their prior mixed educational experiences with laboratory-based anatomy were considered to add undesirable confounds to the study. Studies on medical imaging (e.g., incorporating ultrasound into anatomy), didactic teaching approaches (e.g., TBL, PBL, etc.), and/or supplemental instructional programs (e.g., computer assisted instruction, virtual reality, surgical skills training, etc.) were excluded if they were not laboratory/dissection focused. These topics will be the focus of a future meta-analysis project. As an ad hoc requirement, all studies had to either report empirical data on learner performance or report on the perceptions of learners and/or educators. An intervention in which learners were exposed to laboratory content for some duration of time was required. No geographical restrictions were specified and only studies written in English were included.

**Study Selection and Data Extraction**

The protocol presented herein closely follows a procedure previously described by Wilson et al. (2016). Four teams of paired researchers reviewed roughly one-quarter of all records retrieved. Each member of the team made independent judgments on whether or not to include a record for full-review based on information presented in the records’ titles and abstracts. All discrepancies that arose from each team’s preliminary screening were resolved using a cross-over design. That is, members of a different review team evaluated and made final inclusion/exclusion decisions on records they were not initially responsible for. Decisions regarding these discrepancies often required a full article review and were settled by team consensus. This process further refined the number of studies retained for full-review and ensured agreement regarding the applicability of each study to the goals of the meta-analysis. Cohen’s $\kappa$ statistic and percent agreement were used to calculate inter-rater reliabilities for the dichotomous judgments made between paired researchers (Cohen, 1960). Substantial coding agreement between researchers was considered to be achieved if a Cohen's $\kappa$ statistic was reported to be 0.61 or higher (Landis and Koch, 1977). A percent agreement of $\geq$ 70% was considered to demonstrate adequate agreement (Birkimer and Brown, 1979). The responsibility of extracting the necessary numerical and qualitative data during the full-review phase was shared equally among the researchers. To minimize bias and variability in data collection and interpretation, the lead author re-reviewed and confirmed the appropriateness of all pertinent data that were included for analysis. In instances where studies had incomplete/unusable numerical datasets, attempts were made to acquire the necessary information from the study’s corresponding author. Studies were ultimately excluded if correspondence was unsuccessful.

**Statistical Analyses**

Data were collected using a customized REDCap (Harris et al., 2009) smart form with branching logic (https://redcap.uits.iu.edu/surveys/?s=HLRRLJE9JP) and were exported to Microsoft Excel® for organization and cleaning. Review Manager (RevMan 5.3) computed heterogeneity and effect sizes (reported as standardized mean differences (SMD) using Hedges’ adjusted $g$) and generated forest and funnel plots. A random-effects model was applied to calculate the summary effect size (Borenstein et al., 2010). Inverse variance was used to weight studies as a function of their sample sizes. While it is relatively common for meta-analyses to report multiple effect estimates from a single study population, statistically speaking this is a violation of the independence assumption that underlies the procedures for aggregating data (Petitti, 2000). To avoid violating this assumption, we chose to estimate composite scores across all reported measures (e.g., practical and written examinations) by averaging the reported means and standard deviations. All measures were given equal weight in the composite score calculation which ultimately allowed a single effect size to be calculated per study. If present, pre-existing global outcomes (e.g., final course scores) were used over estimated composite scores for computing effect size estimates. An analysis of moderator effects in the form of study features (i.e., study design, participant type, intervention length, or specimen type) was also performed to determine whether certain study characteristics had a unique influence on the pooled effect sizes.

The magnitude of the summary effect size was interpreted using Cohen’s (1988) recommendations for small ($0.20–0.49$), medium ($0.50–0.79$), and large ($\geq 0.80$) effects. Confidence intervals (CI) were also reported at the 95% confidence level. A Q statistic was used to determine the presence of heterogeneity (Huedo-Medina et al., 2006) and an $I^2$ statistic estimated the amount of variance between studies. Heterogeneity was considered inconsequential if the variance between studies ($I^2$) was $<25\%$ and was considered substantial if $>75\%$ (Higgins et al., 2003). A prediction interval was calculated to provide additional context. A prediction interval is the range within which an effect size is predicted to fall 95% of the time for any given population (Borenstein et al., 2017). Lastly, a funnel plot was visually studied for
symmetry to gauge the likelihood of publication bias (i.e., whether an imbalance existed between studies with positive findings and studies with negative or inconclusive findings) (Duval and Tweedie, 2000).

Records that were reviewed in-full for the meta-analysis were also scanned for survey-based perceptions data. Reported survey items were reviewed in search of data that explored respondents’ preferences toward cadaver or dissection-based learning. The number and proportion of participants that responded to a related survey item on this topic were tallied to better understand global perceptions across studies and learner populations.

**RESULTS**

**Search and Screening Outcomes**

In total, 3,552 records were returned through the electronic search and three additional records were identified by hand searching. A number of duplicates (520) were removed leaving 3,035 records for title and abstract screening. Upon excluding records, 327 articles underwent a full-text review to further determine their eligibility and to extract data as appropriate. The full-review process excluded an additional 300 studies and yielded 27 studies for the meta-analysis. Figure 1 summarizes the progression of the study inclusion/exclusion process. All data extracted for the meta-analysis came from either published full-text articles or dissertations.

Because of the volume of articles retrieved, the preliminary screening was conducted in two phases. Both a percent agreement and Cohen’s $\kappa$ were calculated for each paired research team to measure inter-rater agreement. Table 1 shows the amount of agreement between investigators by team. Across the two phases, the average percent agreement between researchers was 84.9% and average Cohen’s $\kappa$ was found to be moderate at 0.449.

**Summary of Study Characteristics**

Most studies included in the meta-analysis were conducted within the disciplines of anatomy or biology. As such, dissections were performed on either human or animal specimens. Because the process of dissection was the construct of interest in this analysis (opposed to specimen type), studies that were conceptual replications of one another (e.g., veterinary student versus medical student studies) were deemed appropriate for inclusion in the analysis. Study populations were diverse as they included high school and undergraduate students as well as masters and doctoral level learners. Although study designs varied, the majority of studies included a quasi-experimental, nonrandomized approach. The duration of study interventions also varied in length from being less than a semester long to being a semester in length or longer. Most often, studies utilized identical assessment measures to evaluate the effects of interventions on changes in student performance.

**Meta-Analysis of Anatomy Laboratory Pedagogies**

Figure 2 presents the findings of the meta-analysis in a forest plot. Overall, the summary effect size was calculated in the context of 4,435 subjects exposed to traditional forms of dissection and 3,296 subjects exposed to other forms of anatomy laboratory instruction. According to the aggregated outcomes of 27 studies, no effect (SMD = -0.03, [CI = -0.16, 0.10], $P = 0.62$) was detected when comparing short-term knowledge gains achieved through traditional dissection versus those achieved through other laboratory instructional modalities. Figure 2 also presents the overall effect for each of the four subanalyses which included dissection versus prosection, dissection versus digital media, dissection versus models/modeling, and dissection versus a hybrid approach (which involved variable amounts of dissection and prosection combined). No significant effects ($P \geq 0.15$) were detected for any of the four subanalyses.

A significant Q statistic ($P < 0.001$) indicated the presence of heterogeneity across all 27 studies. A total $I^2$ index of 82% suggested the results between studies were inconsistent with the total variation in study estimates being because of heterogeneity rather than sampling error (Fig. 2). By convention, this represents sizable variation between studies. As per the prediction interval, the predicted range of effects was found to be −0.63 to 0.57. Significant heterogeneity ($I^2$ range: 70–85%) was also present within each subgroup analysis (Fig. 2). The relative symmetry of the funnel plot presented in Figure 3 suggests that publication bias of the sampled studies was unlikely.

**Analysis of Moderator Effects**

An analysis of study features was conducted to determine whether study characteristics (i.e., study design, participant type, intervention length, or specimen type) had a unique influence on the pooled effect sizes. All moderator analyses were found to have non-significant effects ($P \geq 0.16$; Fig. 4). This indicates that study characteristics had no influence on the study’s major findings.

**Perceptions Analysis**

After completing the meta-analysis, the 327 records that underwent full-review were re-examined to extract pertinent perceptions data. Survey data were compiled from 17 studies and involved a total of 3,471 participants. The analysis of student and educator perceptions revealed that, on average, roughly three out of every five respondents (57%) favored or valued cadaver/dissection-based learning over other laboratory modalities (Table 2).

**DISCUSSION**

The literature is rich with debates regarding the role of dissection in educational programs. Many have lamented that too few studies have fully compared
the efficacy of traditional dissection programs to the instructional approaches which replace them (Pawlina and Lachman, 2004; Rizzolo and Stewart, 2006; Winkelmann, 2007; Sugand et al., 2010). The recent and explosive growth of educational literature dedicated to the anatomical sciences has kindled our desire to confront the dissection debate head-on by viewing the extant evidence through the perspective of a meta-analysis. This meta-analysis revealed that speculations drawn from prior thematic reviews do not accurately depict the breadth of available evidence. Contrary to the notion that dissection shows a slight advantage over prosection (Winkelmann, 2007), our findings suggest there are no immediate benefits of implementing dissection over other instructional modalities in terms of learner achievement, as
measured through short-term knowledge-based examinations.

Interestingly, historical evidence suggests a profound role for dissection in health professions education. This is evidenced, in part, by the reversal of the decade old “trend” to discontinue cadaveric dissection. Some American medical schools that once abandoned dissection (e.g., New York University, University of California at San Francisco and Davis, University of Hawaii, and University of Washington) have now reverted back to practicing the time honored tradition (Rizzolo and Stewart, 2006). The trend to restore anatomical dissection was also noted by researchers who examined the widespread use of prosections in Australia and New Zealand (Craig et al., 2010). This is despite a belief, “that a well-designed anatomical curriculum without cadaver dissection might be as good as or even better than cadaver dissection in the learning of gross anatomy (McLachlan et al., 2004)”. The about-face and eagerness of some schools to return to dissection was, thusly, thought to be provoked by undesirable learner performance and driven by student preferences for donor dissection; irrespective of the cost savings schools had grown accustomed to (Rizzolo and Stewart, 2006). This theory, however, is largely anecdotal and is not fully supported by our findings. As our study reveals, the use of nondissection modalities was of no detriment to students in regards to their short-term academic performance. However, student preference for anatomic dissection may be unrelated to student achievement outcomes and more related to students’ confidence levels and their ability to practically and competently demonstrate their knowledge of anatomy in clinical settings. This begs the question, ”What pedagogical approach maximizes learners’ confidence and their ability to clinically problem solve in the future?” Students’ preferences for dissection may also be attributed to their desire for new experiences and a yearning to learn new skillsets as Larkin and McAndrew (2013) have proposed.

Thus the central question lingers, why does donor dissection remain a pivotal aspect of traditional and modern healthcare professions curricula? Could there be something inherent to the rhythm and process of dissection that ultimately strengthens the attributes or deepens the knowledge-base of clinically oriented trainees? One answer may be the ability of donor dissection to unwittingly impart skills imperative to the healthcare professions (i.e., teamwork, surgical skills, etc.). Another explanation may be that the process of discovering and presenting anatomical structures parallels diagnostic reasoning, the underpinnings of which are built on the tenants of observation, interpretation, and differential identification (Rizzolo and Stewart, 2006). While there is literature supporting the role for dissection in developing professional skills (Pawlina and Lachman, 2004; Escobar-Poni and Poni, 2006; Lachman and Pawlina, 2006), there remains a dearth of studies exploring the role of dissection alternatives in imparting such skills. Is the acquisition of professional skills inherent solely in donor dissection activities or is it perhaps integrated more holistically in the broader fabric of a well-rounded anatomy education curriculum?

Another aspect our meta-analysis revealed as lacking in the current literature was the effect of the investigated pedagogies on the long-term retention and application of anatomical knowledge in clinical settings. While we observed no immediate differences in student performance between instructional modalities, it is plausible for schools that initially abolished cadaveric dissection to have incidentally observed a marked decrease in learning outcomes or technical performance at other time points in their medical curricula (e.g., clinical years), as purported in a commentary by Rizzolo and Stewart (2006). The oft-reported perception that medical school trainees graduate with an inadequate and unsafe level of anatomical knowledge further supports the theory that preclerkship curricular decisions may in fact have more lasting effects than initially realized (Turney, 2007). Unfortunately, little empirical evidence substantiates the premise that an absence of dissection causally influences the poor retention of anatomical knowledge. In spite of the paucity of literature, Custers’ (2010) educational review informs us that the most important determinant of long-term retention (that is under an educator’s or administration’s control) is prolonged contact with the content under study. Perhaps the relative benefits of time intensive dissection are not as evident until learners have been far removed from the domain of study.

Is the reason for maintaining donor dissection because of its inherent “active learning” qualities? Educational evidence suggests that active learning and student engagement lead to significant improvements in the recall of information because, philosophically, direct experiential opportunities are necessary for cultivating deep learning (Prince, 2004; Cake,

---

### TABLE 1. Summary of Investigator Agreement for Abstract/Title Screening

<table>
<thead>
<tr>
<th>Team</th>
<th>Percent agreement</th>
<th>Cohen’s $\kappa$</th>
<th>No. of records screened</th>
<th>Percent agreement</th>
<th>Cohen’s $\kappa$</th>
<th>No. of records screened</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team 1</td>
<td>80.0%</td>
<td>0.279</td>
<td>305</td>
<td>91.4%</td>
<td>0.450</td>
<td>533</td>
</tr>
<tr>
<td>Team 2</td>
<td>67.4%</td>
<td>0.289</td>
<td>304</td>
<td>88.0%</td>
<td>0.461</td>
<td>533</td>
</tr>
<tr>
<td>Team 3</td>
<td>80.8%</td>
<td>0.476</td>
<td>302</td>
<td>93.1%</td>
<td>0.609</td>
<td>533</td>
</tr>
<tr>
<td>Team 4</td>
<td>N/A</td>
<td>N/A</td>
<td>0</td>
<td>93.9%</td>
<td>0.577</td>
<td>522</td>
</tr>
<tr>
<td>Mean [Total]</td>
<td>76.1%</td>
<td>0.348</td>
<td>[911]</td>
<td>91.6%</td>
<td>0.524</td>
<td>[2121*]</td>
</tr>
</tbody>
</table>

*aThe three articles retrieved via hand-searching were identified during the full-review phase of the study, hence the discrepancy between 3,035 total records reported in Figure 1 and 3,032 total records reported here.*
Fig. 2. Random-effects model forest plot with summary and study specific effects. SMD: standardized mean difference; SE: standard error. [Color figure can be viewed at wileyonlinelibrary.com]
This, in part, may explain why the schools that once abandoned dissection, in general, observed poor anatomy related performance during their students' clinical years after having originally dismissed dissection as an approach of the past. However, merely implementing a dissection-based curriculum does not always directly correlate to active participation in anatomy learning (Winkelmann et al., 2007). More research is needed to better understand the temporal consequences of replacing cadaveric dissection with more modern teaching approaches.

The debate regarding the centrality of dissection to anatomy education is further complimented by national data and trends. A recent commentary published through the Association of American Medical Colleges (AAMC) presented trends on "Resources for Learning Anatomy" (Mintz, 2016). Of the 134 schools that provided information on gross anatomy, 132 (99%) indicated the use of cadaver dissection and 119 (89%) indicated the use of projections. This suggests to the authors that a large majority of US and Canadian medical schools are using a hybrid/multifaceted approach to laboratory instruction. The use of plastinated specimens (68 of 134; 51%), anatomical models/simulators (104 of 134; 78%), and virtual/online anatomical manuals (109 of 134; 81%) was also found to be common place among anatomy programs at medical institutions. Based on our findings, and the general lack of superiority of any one instructional method, we posit that anatomy educators find value in the unique benefits that each instructional approach has to offer. While a thorough discussion of the advantages and disadvantages of laboratory approaches is beyond the scope of this study, we contend that a multifaceted model of anatomy education has the greatest chance of fulfilling the needs of diverse learner populations. It is a middle ground that gives direct hands-on experience to the slight majority who favor/value dissection while offering oppositionists of dissection an outlet for less physical engagement than may otherwise be required.

In many ways cadaveric dissection can be viewed as a form of simulation, as cadavers have both physical and functional fidelity (Maran and Glavin, 2003). Just as simulators are said to be only as good as the educational curriculum within which they are embedded, we believe this principle also applies to dissection programs (Maran and Glavin, 2003). Nnodim (1990) remarked that, "the effectiveness of a learning method might have more to do with the way that method is applied in the given circumstances than with the method itself." In light of our nonsignificant findings, we believe that the educational effectiveness of cadaveric dissection and any other instructional approach employed in the anatomy laboratory is predicated upon meaningful and purposeful interactions between the students, the approach(es) chosen, and the
Moreover, learning tends to be optimized when instructional strategies are deliberate and delivered within the boundaries of predefined objectives.

Limitations

In general, the outcomes of meta-analyses are subject to the comparability of different samples and are restricted by the dissimilarities of research designs and outcome measures (Davies, 1999). These factors, in addition to the amount of conflicting data, are likely to explain why the observed heterogeneity was high ($I^2=82\%$, $P<0.001$). Because the outcome of the meta-analysis failed to reject the null hypothesis, a retrospective power analysis was conducted, according to formulas proposed by Valentine et al. (2010), to determine whether there was sufficient power to detect the smallest meaningful effect. A two-tailed power analysis was computed using the smallest meaningful effect of 0.20, the variance of the observed summary effect size ($v=0.10$), and the standard normal cumulative distribution function ($\Phi(-0.02)$) at alpha = 0.05. Given the number of studies and weights assigned to each study (a function of the within-study sample sizes), power to detect what we believe to be the smallest important effect was moderate ($P = 0.538$).

Future Directions

While this study clarifies aspects of the dissection debate, additional research is needed in three key areas. First, we propose a "sister" meta-analysis be conducted to explore the effects of more modern didactic pedagogies (e.g., team-based learning, instructional strategies are deliberate and delivered within the boundaries of predefined objectives.

### TABLE 2. Proportion of Survey Participants Favoring/Valuing Cadaver or Dissection-Based Learning

<table>
<thead>
<tr>
<th>Study</th>
<th>Number (%) favoring/valuing cadaver or dissection based learning</th>
<th>Survey item/Contextual notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Sinclair, 1965)*</td>
<td>29 of 92 (32%)</td>
<td>Evaluated the use of prosections as a lab approach: Better than, as good as, or inferior to (reported percentage) dissection</td>
</tr>
<tr>
<td>(Stillman et al., 1978)</td>
<td>22 of 88 (25%)</td>
<td>How did using live models compare with using cadavers in helping you understand the relationship between superficial and deep anatomical structures?</td>
</tr>
<tr>
<td>(Nnodim, 1990)b</td>
<td>15 of 21 (71%)c</td>
<td>Learning from prosection should completely replace learning by dissection</td>
</tr>
<tr>
<td>(Berube et al., 1999)</td>
<td>95 of 110 (86%)</td>
<td>Are cadavers a necessary component in anatomy instruction?</td>
</tr>
<tr>
<td>(Predavec, 2001)b</td>
<td>35 of 168 (21%)</td>
<td>Would a real dissection (as opposed to E-Rat) be a better educational experience?</td>
</tr>
<tr>
<td>(Leung et al., 2006)</td>
<td>90 of 94 (96%)</td>
<td>Cadaver dissection: Very helpful, helpful, neutral, unhelpful, very unhelpful</td>
</tr>
<tr>
<td>(Petersson et al., 2009)*</td>
<td>19 of 35 (54%)</td>
<td>How does learning of vascular anatomy with [digital media] compare to dissections/autopsies?</td>
</tr>
<tr>
<td>(Ahmed et al., 2010)*</td>
<td>141 of 222 (64%)</td>
<td>Select the learning tool which [you] most recommend for learning basic anatomy.</td>
</tr>
<tr>
<td>(DeHoff et al., 2011)*b</td>
<td>24 of 74 (32%)</td>
<td>I prefer using the experimental method [dissection] as the primary method of laboratory instruction for learning human anatomy</td>
</tr>
<tr>
<td>(Fruhstorfer et al., 2011)</td>
<td>82 of 124 (66%)</td>
<td>I would have preferred to have dissection experience in addition to plastinated prosections.</td>
</tr>
<tr>
<td>(Zurada et al., 2011)*</td>
<td>444 of 705 (63%)</td>
<td>Preferences towards using cadavers and radiological images for the purpose of understanding topography and spatial relationships of various structures.</td>
</tr>
<tr>
<td>(Duran et al., 2012)</td>
<td>195 of 558 (35%)</td>
<td>Value placed on the practice of dissection and prosection in the laboratory.</td>
</tr>
<tr>
<td>(Mulu and Tegabu, 2012)</td>
<td>136 of 147 (93%)</td>
<td>Which method do you prefer for anatomy learning? Dissection or Prosection</td>
</tr>
<tr>
<td>(Mustafa et al., 2013)</td>
<td>124 of 246 (50%)</td>
<td>I prefer to use cadavers and/or prosections. Would you have preferred alternate methods other than dissecting the rat to learn human anatomy?</td>
</tr>
<tr>
<td>(Haspel et al., 2014)*</td>
<td>27 of 71 (38%)</td>
<td>I prefer to use cadavers and/or prosections. Would you have preferred alternate methods other than dissecting the rat to learn human anatomy?</td>
</tr>
<tr>
<td>(Schofield, 2014)</td>
<td>43 of 50 (86%)</td>
<td>Recommendation for anatomy course in OT curricula; inclusion of cadaver dissection and/or prosection</td>
</tr>
<tr>
<td>(Marom and Tarrasch, 2015)</td>
<td>457 of 666 (69%)</td>
<td>Anatomy teaching should be dissection-based (versus image-based).</td>
</tr>
</tbody>
</table>

**TOTAL**

1,978 of 3,471 (57%)

Studies with Likert-scale data presented as means and standard deviations were excluded from the analysis because data in this format could not be meaningfully aggregated to understand the prevalence of perceptions.

*aData were combined from multiple participant populations

*bStudy included in meta-analysis;

*cInverse data (e.g., "Disagree" data) were reported for negatively worded survey items.

Moreover, learning tends to be optimized when instructional strategies are deliberate and delivered within the boundaries of predefined objectives.

Limitations

In general, the outcomes of meta-analyses are subject to the comparability of different samples and are restricted by the dissimilarities of research designs and outcome measures (Davies, 1999). These factors, in addition to the amount of conflicting data, are likely to explain why the observed heterogeneity was high ($I^2=82\%$, $P<0.001$). Because the outcome of the meta-analysis failed to reject the null hypothesis, a retrospective power analysis was conducted, according to formulas proposed by Valentine et al. (2010), to determine whether there was sufficient power to detect the smallest meaningful effect. A two-tailed power analysis was computed using the smallest meaningful effect of 0.20, the variance of the observed summary effect size ($v=0.10$), and the standard normal cumulative distribution function ($\Phi(-0.02)$) at alpha = 0.05. Given the number of studies and weights assigned to each study (a function of the within-study sample sizes), power to detect what we believe to be the smallest important effect was moderate ($P = 0.538$).

Future Directions

While this study clarifies aspects of the dissection debate, additional research is needed in three key areas. First, we propose a "sister" meta-analysis be conducted to explore the effects of more modern didactic pedagogies (e.g., team-based learning,
problem-based learning, case-based learning, computer-assisted learning, etc.) on knowledge gains in anatomy. While a study by Freeman et al. (2014) suggests that pedagogies with active learning components enhance student performance across disciplines and student populations in higher education, there is still a need to study the effects of these approaches on the acquisition of anatomical knowledge specifically. Secondly, the anatomy education literature could benefit from systematically investigating the hidden benefits of dissection. The process of dissection, and all it entails, could have profound ripple like effects that influence other areas of learning and identity formation in students. Lastly, the effects of certain pedagogical approaches on the long-term retention of anatomical knowledge are not well understood, and as such research in this area is recommended.

CONCLUSIONS

The anatomy education literature is rife with extensive discussions of the advantages and disadvantages of dissection and commentaries on the opportunities dissection offers for psychosocial and professionalism development. The contribution of this work focused on comparing the effectiveness of dissection to other laboratory instructional approaches. Student performance scores were found to be statistically equivalent when comparing traditional dissection to other laboratory methods (i.e., prossection, digital media, 3 D models/ modelling, hybrid approaches). We therefore encourage educators to critically dwell on the intentionality and application of their selected laboratory approach(es) as this aspect of the educational design process may carry more weight in fostering student success than the innate attributes of the chosen method itself. Without more substantial evidence that explains how the ripple effect of dissection may impact other areas of medical and health professions curricula, the authors are hesitant to declare the dissection debate resolved. While much progress has been made, work remains to bring the scope of this debate into even clearer focus.

ACKNOWLEDGMENTS

The authors would like to extend a warm thank you to Dr. Mark Seifert of Indiana University School of Medicine for reviewing and providing feedback on an early draft of this manuscript.

ETHICAL APPROVAL

This study did not involve human subjects and as such was not subject to IRB approval.

PRESENTATIONS

This study was presented at the American Association of Anatomists 2017 Experimental Biology conference. FASEB Journal, vol. 31. no. 1 Supplement 392.1

REFERENCES


