



Широбоков И.Г.¹⁾, Пугачева Е.В.²⁾

¹⁾ МАЭ РАН, Отдел антропологии, Университетская наб.,
3, Санкт-Петербург, 199034, Россия

²⁾ Высшая нормальная школа (Париж), 45, rue d'Ulm / 29 rue d'Ulm /
24 rue Lhomond, F-75230 Paris cedex 05, Франция

О ТОЧНОСТИ ВИЗУАЛЬНОЙ ОЦЕНКИ ВОЗРАСТА ПО ЧЕРЕПУ ВЗРОСЛОГО ЧЕЛОВЕКА (МЕТОДИЧЕСКИЕ АСПЕКТЫ)

Существует множество формализованных подходов к оценке возраста по черепу, большинство из которых опирается на анализ закономерностей в степени облитерации швов и стертости зубов. Однако многие отечественные антропологи оценивают возраст в произвольных интервалах, не фиксируя признаки в конкретной системе и не определяя их относительную значимость. Задачи настоящего исследования состоят в анализе эффективности субъективно-визуального подхода при оценке индивидуального возраста и реконструкции возрастной структуры выборки.

Материал и методы. Исследование основано на анализе серии 116 черепов начала XX века с задокументированным полом и возрастом из фондов МАЭ РАН. Два автора независимо друг от друга оценивали возраст, а также фиксировали степень облитерации швов свода черепа и степень стертости жевательной поверхности зубов. Корреляция между возрастом и балловым признаком оценивалась при помощи коэффициента корреляции Спирмена, расхождения между реальным и прогнозируемым возрастом – при помощи средней абсолютной ошибки и систематической ошибки. Расчеты проводились как для всей выборки, так и для отдельных возрастных групп. Согласованность оценок между исследователями оценивалась при помощи коэффициента внутри-классовой корреляции.

Результаты. Оценки авторов обнаружили умеренно высокую согласованность между собой и умеренную положительную корреляцию с реальным возрастом. Точность визуальных оценок не уступает точности формализованных методов, основанных на анализе степени облитерации швов. Все оценки демонстрируют эффект регрессии к среднему: возраст индивидов в младших когортах систематически завышается, возраст старших когорт, напротив, занижается. Точность определения возрастной структуры выборки в некоторой степени зависит от структуры выборки. Усреднение оценок разных авторов или оценок одного автора, повторно установленных с большим временным интервалом, позволяет приблизить оценки к реальной картине.

Заключение. Увеличение межисследовательской сопоставимости оценок возраста может быть достигнуто за счет фиксации признаков в одинаковых балловых шкалах, а также за счет увеличения возрастных интервалов. Точность оценок может быть повышена за счет проведения повторного исследования возраста, а также эффекта «мудрости толпы».

Ключевые слова: швы свода черепа; оценка возраста; возрастная структура выборки; регрессия к среднему; мудрость толпы

DOI: 10.55959/MSU2074-8132-24-4-8

Введение

Морфологическая, половая и возрастная изменчивость признаков черепа традиционно является одним из наиболее популярных предметов исследований антропологов, специализирующихся на работе со скелетными останками. Исторически так сложилось, что антропологи придавали особое значение именно изучению черепа, и это имело важные долгосрочные последствия для нашей науки, в т.ч. и на современном этапе ее развития. Значительную часть антропологических коллекций, поступавших в музеи до середины 20 века, представляют исключительно краниологические серии, тогда как кости посткраниального скелета составляют в них заметно меньшую долю. Череп зачастую является единственным источником информации для получения как индивидуальной, так и групповой характеристики древнего населения, и качество такой информации не всегда можно считать удовлетворительным. По сравнению с комплексными методиками определения возраста оценки, вынесенные по отдельным элементам скелета, имеют заметно более ограниченную точность. Подходы к установлению биологического возраста скелета (а вместе с ним и величине допустимого интервала в хронологической шкале) отличаются заметной вариативностью, и вопрос о сопоставимости таких оценок и возможных методах повышения их точности, в т.ч. при использовании исследователями недостаточно строго формализованных методик, остается в числе вечно актуальных тем антропологии.

Один из наиболее ранних и по-прежнему широко используемых методов определения возраста скелета взрослого человека основывается на оценках степени облитерации швов свода черепа [Todd, Lyon, 1924; Acsadi, Nemeskeri, 1970; Meindl, Lovejoy, 1985; Key et al., 1994; и др.]. Темпы этого процесса различаются в разных отделах черепа и зависят от множества внутренних и внешних факторов, в т.ч. от пола, формы черепа, генетических особенностей, гормонального статуса, диеты, биомеханических нагрузок и прижизненной деформации [Никитюк 1960; Звягин 1976; Wolff et al., 2013; Ruengdit et al., 2020; и др.].

Высокая индивидуальная изменчивость скорости облитерации швов ограничивает возможности самостоятельного использования этой характеристики при решении задач судебно-медицинской экспертизы, и все же соответствующие методы остаются востребованными в

среде судебных антропологов [Ruengdit et al., 2020]. Кроме того, методы пользуются неизменной популярностью среди антропологов, работающих с материалами из археологических раскопок. В этом случае не столь важна точность индивидуальных оценок возраста (определения в широких категориях вполне достаточны), сколько возможность использования метода при установлении возрастной структуры выборки. Другой возрастной характеристикой, наиболее часто используемой при оценке возраста скелета как на индивидуальном, так и групповом уровнях, остается степень стертости жевательной поверхности зубов. Как показывают данные обзора статей, опубликованных в журналах «American Journal of Physical Anthropology», «International Journal of Osteoarchaeology» и «Journal of Archaeological Science», наибольшей популярностью пользуется схема Д. Бротвелла, за которыми следуют методы, предложенные А. Майлзом и О. Лавджоем [Falys, Lewis, 2010]. Отечественные антропологи, по-видимому, пользуются преимущественно схемой М.М. Герасимова, основанной на видоизмененной шкале П. Брока [Герасимов, 1955], хотя, следует признать, что точный анализ популярности методик не проводился. Так или иначе, очевидно, что степень стертости зубов также зависит от множества факторов (пищи, прижизненной утраты зубов), и ни одна универсальная схема не может претендовать на роль действительно эффективной.

Поскольку существует несколько различных методов оценки возраста, основанных на анализе признаков черепа, стороннему читателю может показаться удивительным тот факт, что из публикаций антропологов, как правило, невозможно установить какой же именно метод был использован. Обычно авторы ограничиваются ссылками на некоторый список работ, в которых описаны разные подходы к оценке. Нередко также встречаются тексты, в которых просто указывается, что при определении возраста учитывались степень заращения швов и степень стертости зубов. Вероятно, причина такой практики заключается том, что в действительности исследователи чаще всего ориентируются на субъективную визуальную оценку возраста, не прибегая к строго формализованным методам. Мы обращаем внимание на те признаки возрастных изменений черепа, сведения о диагностической ценности которых почерпнули из литературы и практического опыта исследований,

но, как правило, не фиксируем их в измерительных шкалах и не проводим строгий анализ их относительной значимости при вынесении итоговой оценки. Величина возрастных интервалов также определяется произвольно – либо в виде 5- или 10-летнего интервала, либо в формате возрастной категории (*adultus*, *maturus* и др.).

Эта проблема не является специфической для российской антропологии. Например, как показывает обзорное исследование, авторы большинства статей, опубликованных в журналах «*American Journal of Physical Anthropology*», «*International Journal of Osteoarchaeology*» и «*Journal of Archaeological Science*» в 2004–2009 гг., приводят ссылки на методические работы, но не дают описания конкретных использованных методов и подхода, использованного для вынесения окончательной оценки. В некоторых публикациях методы оценки возраста вообще не упоминаются, а в большинстве остальных случаев указывается более одного метода [Falys, Lewis, 2010].

Цель настоящего исследования заключается в том, чтобы проанализировать точность субъективно-визуального подхода к оценке возраста по черепу по сравнению с формализованными методами, основанными на анализе степени заращения швов свода черепа, а также изучить влияние систематических ошибок в определениях возраста на надежность оценки возрастной структуры выборки. В обсуждении рассматриваются перспективы применения некоторых подходов, которые могут быть использованы для повышения точности оценок возраста и их межисследовательской согласованности.

Материалы и методы

Материалами исследования послужили черепа из коллекций с задокументированным полом и возрастом смерти в антропологических фондах МАЭ РАН. Черепа были переданы в Кунсткамеру в 1910-х гг. К.З. Яцутой из собрания Военно-Медицинской Академии. В МАЭ РАН они зарегистрированы в составе коллекций 1830, 1831, 1989, 1994, 2222, 2223, 2282 и 2303. В коллекциях представлены черепа преимущественно русских из различных регионов европейской России, а также (в единичных случаях) немцев, финнов и поляков. Всего в исследовании были учтены информация о возрастной изменчивости 116 черепов, большинство из кото-

рых принадлежало мужчинам. Возраст умерших колеблется от 18 до 70 лет.

Возраст оценивался следующим образом.

Сначала авторы оценивали приближенные возрастные интервалы, в пределах которых, по их мнению, находится истинный возраст индивидов, ориентируясь на макроскопические признаки возрастных изменений черепа, но не фиксируя их значения в каких-либо стандартизованных шкалах. Учитывались степень заращения швов свода черепа как на внешней поверхности, так и со стороны эндокрана, степень стертости жевательной поверхности зубов, признаки атрофии альвеолярного отростка и тела нижней челюсти, относительный вес и плотность костей черепа.

Затем проводилась оценка степени стертости зубов в соответствии со схемой М.М. Герасимова [Герасимов, 1955],¹ а также определялась степень заращения венечного, сагиттального и лямбдовидного швов черепа. Использовалась четырех-балльная система: 0 – шов полностью открыт, 1 – закрыто менее половины шва, 2 – закрыта половина шва или более, 3 – полная облитерация шва. Степень облитерации оценивалась как со стороны эндокрана, так и на внешней поверхности. Для каждого черепа проводился расчет коэффициентов облитерации швов (КОШ) – суммы всех баллов, а также сумм баллов, рассчитанных для эндокрана и внешней поверхности отдельно.

Большинство существующих методик предполагает, что оценка степени облитерации швов производится на отдельных участках в 4- или 5-балльной системе. Однако поскольку хорошо известно, что корреляции между степенью облитерации швов и паспортным возрастом человека, как правило, невелики, авторы использовали упрощенный вариант оценки. Ее сравнительная эффективность оценивалась следующим образом. Оценки, полученные первым автором, были использованы для обучения модели линейного дискриминантного анализа. При этом каждый череп поочередно выступал в качестве тестового, а остальная часть – в качестве обучающей выборки. Расхождения между реальным и прогнозируемым возрастом оценивались при помощи средней абсолютной ошибки

¹ На большинстве черепов вследствие прижизненной и посмертной утраты зубов отсутствует большая часть зубного ряда, у многих отсутствуют нижние челюсти, поэтому возможность использования оценок стертости зубов была заметно ограничена.

(MAE) и систематической ошибки (CO). Последняя рассчитывалась как средняя разница между установленным и задокументированным возрастом для отдельных возрастных групп (10-летних когорт). Полученные результаты сопоставлялись с результатами оценки двух популярных методик, основанных на анализе возрастной изменчивости швов свода черепа – Д. Ачади и Я. Немешкери [Acsadi, Nemeskeri, 1970] для эндокрана, а также Р. Меиндла и О. Лавджоя [Meindl, Lovejoy, 1985] для внешней поверхности. При сопоставлении оценок привлекались литературные данные.

Корреляции между возрастом и КОШ, возрастом и средним баллом стертости зубов рассчитывались при помощи коэффициента корреляции Спирмена (r_s). Согласованность оценок возраста и КОШ между исследователями оценивалась при помощи коэффициента внутриклассовой корреляции (ICC). В отличие от коэффициента корреляции Пирсона ICC позволяет учитывать влияние систематических расхождений между исследователями. Коэффициент учитывает соотношение двух разных компонентов общей дисперсий в оценках: дисперсию, обусловленную различиями между оцениваемыми объектами, и дисперсию, обусловленную различиями между оценщиками или методиками. Существует несколько типов коэффициента внутриклассовой корреляции, предназначенных для оценки согласованности при разных условиях задачи. При оценке согласованности результатов между авторами использовался ICC (2, k), между оценками каждого из авторов и задокументированными данными – ICC (3, k) (обозначения приведены согласно классификации [Shrout, Fleiss, 1979]).

До завершения оценки авторам была неизвестна ни возрастная структура выборки, ни крайние значения возраста индивидов. Более того, коллекционные номера черепов в рабочих файлах зашифрованы таким образом, чтобы даже по завершении представленного исследования у авторов не было возможности оценить вероятные причины ошибок в оценках возраста, повторно обратившись к изучению конкретных черепов. Такие меры были приняты для более корректного проведения возможных дополнительных исследований в дальнейшем. В представленном виде статья намечает наиболее важные шаги для дальнейших исследований, которые будут опираться на более представительный материал и которые предполагают расширение числа участников.

Результаты

В целом оценки возраста (учитывались средние оценки, рассчитанные для конкретных интервалов) обнаружили умеренно высокую степень согласованности между авторами. Внутриклассовая корреляция между авторами как по оценкам возраста, так и по величине КОШ оказалась одинаковой (ICC (2, k)=0,79), корреляция между авторами по оценкам средних баллов степени стертости зубов – наиболее высокой (ICC (2, k)=0,89). При разбивке выборки на две равные возрастные группы (до 40 лет и старше) обнаружилось, что корреляции между авторами выше при оценке признаков, чем непосредственных оценок возраста (табл. 1).

Таблица 1. Корреляции между оценками авторов (ICC)
Table 1. Correlation between observers estimates (ICC)

Возрастной состав	Оценка возраста	Коэффициент облитерации швов	Средний балл стертости зубов
Вся выборка	0.79 [0.70-0.86]	0.79 [0.52-0.89]	0.89 [0.74-0.94]
До 40 лет	0.70 [0.49-0.82]	0.83 [0.71-0.90]	0.81 [0.59-0.90]
Старше 40 лет	0.73 [0.53-0.84]	0.71 [-0.06-0.89]	0.92 [0.75-0.97]

Примечания. В скобках приведен 95%-ый доверительный интервал для оценки ICC.

Notes. The 95% confidence interval for the ICC assessment is shown in parentheses.

Внутриклассовые коэффициенты корреляции (ICC(3, k)) между реальным и установленным возрастом составили 0,72 и 0,70 по оценкам первого и второго авторов соответственно, т.е. достаточно близки к уровню согласованности оценок между исследователями (рис. 1). Средняя абсолютная ошибка составила 9,8 и 8,6 года соответственно, при этом в обоих случаях систематическая ошибка оказалась близкой к нулю (CO=0,1 и 1,1 года). Как будет показано ниже, информация о точности оценок возраста в конкретных возрастных группах заставляет внести существенные коррективы в этот результат.

В большинстве случаев оценки возраста были представлены авторами в виде 10-летних интервалов, реже 5- или 15-летних, а для старших возрастных групп использовались открытые интервалы (старше 50 лет, старше 55 лет). За-

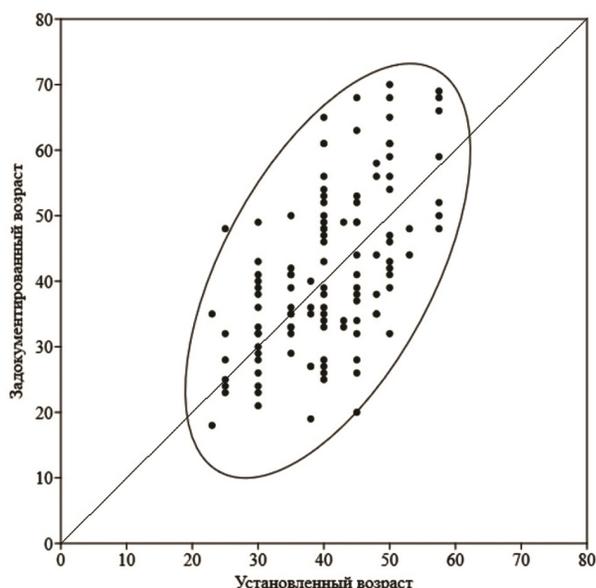


Рисунок 1. Корреляция между установленными и задокументированными оценками возраста (по данным первого автора)

Figure 1. Correlation between estimated and documented age (according to the first author)

документированные оценки возраста оказались в пределах границ интервалов в 40% случаев у первого автора и в 33% у второго. Среднее отклонение задокументированного возраста от границ интервала составило 8,4 и 6,4 года соответственно (учитывались только случаи, в которых возраст не попал в установленный интервал), при этом в 17% случаев у первого автора и в 38% случаев у второго отклонение составило не более 3 лет. Это означает, что переход от 10-летних к 15-летним интервалам позволил бы увеличить число случаев, в которых возраст оказался в границах исследовательских оценок, в полтора-два раза (до 57–61%).

Возможность анализа состояния зубной системы никак не повлияла на точность исследовательских оценок возраста. Об этом свидетельствуют расчеты, проведенные по данным первого автора. Корреляция между МАЕ и комплектностью зубов (суммой сохранившихся зубов и зубов, утраченных при жизни) близка к нулю ($r_s=0,04$). Во многом этот результат обусловлен тем, что прижизненная утрата зубов, фиксируемая даже у индивидов относительно молодого возраста, серьезно повлияла на скорость стирания жевательной поверхности – в одних случаях это привело к увеличению нагрузки, в других (при отсутствии антагонистов), напротив, к ее заметному снижению. Поскольку у многих черепов отсутствуют нижние челюсти, оценить вли-

яние прижизненных утрат на сохранившиеся зубы не всегда представлялось возможным.

Неудивительно, что корреляции между средним баллом стертости зубов и задокументированным возрастом оказались заметно ниже аналогичного показателя, рассчитанного для коэффициентов облитерации швов ($r_s=0,22-24$ и $0,39-0,43$ соответственно). Этот результат показывает, что авторы выбрали верную стратегию, придавая наименьшее значение степени стертости зубов и преимущественно ориентируясь на степень облитерации швов и общее впечатление о признаках возрастных изменений костной структуры (утоньшение костей свода, уменьшение плотности, атрофические изменения). По данным первого автора коэффициенты облитерации швов со стороны эндокрана коррелируют с задокументированным возрастом сильнее, чем коэффициенты, рассчитанные для внешней поверхности ($r_s=0,45$ и $0,36$, $p<0,001$ соответственно). То же самое утверждение справедливо для корреляций КОШ с возрастом, установленным авторами, – в этом случае выше также общая величина коэффициентов корреляции ($r_s=0,69$ и $0,57$, $p<0,001$ соответственно).

В таблице 2 приведены значения средней абсолютной разницы и систематической ошибки между оценками исследователей и задокументированным возрастом для отдельных возрастных интервалов. Разброс оценок (о которой мы можем судить по величине MAE) имеет примерно одинаковую величину в разных группах (и минимальна у автора 2 в интервале 30–49 лет). Фактически значение погрешности заметно возрастает у обоих авторов только в интервале 60–70 лет. При этом возраст индивидов 18–29 лет систематически завышается обоими авторами, а возраст индивидов старше 50 лет, напротив, занижается.

По сравнению с результатами тестирования двух классических методов оценки возраста по степени облитерации швов визуальный подход не обнаруживает заметных недостатков (табл. 3). По сравнению с методом оценки облитерации швов со стороны эндокрана, предложенным Д. Ачади и Я. Немешкери [Acsadi, Nemeskeri, 1970], визуальный подход позволяет получить меньшие или сопоставимые средние отклонения для возрастных групп 18–59 лет и большие отклонения в группе старше 60 лет. По сравнению с методом Р. Меиндла и О. Лавджоя для свода черепа [Meindl, Lovejoy, 1985] визуальный подход в целом менее точен для оценки

всех возрастных групп, за исключением интервала 18–29 лет (для которого точность подходов сопоставима) и интервала 50–59 лет (который точнее позволяет определить именно визуальный подход). При этом по величине систематической ошибки визуальные оценки в целом выглядят более точными по сравнению с обеими методиками. Модель дискриминантного анализа, обученная на упрощенной системе суммарной

оценки степени облитерации швов, также не обнаруживает серьезных недостатков по сравнению с указанными двумя методиками. Средняя величина ошибки для различных возрастных групп в целом сопоставима с таковой для оценок, полученных в рамках методики Р. Меиндла и О. Лавд-жоя, и ни в одном случае не выходит за пределы точности, рассчитанной для обеих методик.

Таблица 2. Величина средней абсолютной разницы и систематической ошибки между оценками и задокументированным возрастом в различных возрастных группах
Table 2. Bias and mean absolute difference between estimated and documented age in different age groups

Возрастная группа, лет	MAE, лет		Систематическая ошибка, лет		Доля корректных интервальных оценок, %	
	1	2	1	2	1	2
Авторы						
18–29	9,1	9,1	7	9	45,5	31,8
30–39	9,6	6,0	7	3	36,1	48,6
40–49	9,9	7,5	0	–3	37,9	31,0
50–59	8	9,5	–2 (–5)*	–7 (–10)*	42,9	12,5
60–70	14,5	15,8	–15 (–17)*	–16 (–18)*	38,5	23,1

Примечания. В скобках приведены расчеты только для тех индивидов, оценка возраста которых авторами была дана в форме закрытого интервала.

Notes. Calculations are given in parentheses only for those individuals whose age was estimated by the authors in the form of a closed interval.

Таблица 3. Средняя абсолютная разница и систематическая ошибка между установленным и задокументированным возрастом в различных возрастных группах по литературным данным

Table 3. Bias and mean absolute difference between estimated and documented ages in different age groups according to literature data

Возрастная группа		Упрощенная 4-балльная система оценки	Внешняя поверхность свода [Meindl, Lovejoy, 1985]			Эндокран [Acsadi, Nemeskeri, 1970]		
			Настоящее исследование (дискриминантный анализ)	Wolff et al. 2012	Ruengdit et al. 2018	Galera et al. 1998	Wolff et al. 2012	Ruengdit et al. 2018
≤29	MAE	11,0	9,0	8,5	12,7	11,6	0,6	12,7
	CO	9,9	9,0	8,5	12,7	10,9	0,6	12,4
30–39	MAE	9,8	4,4	5,7	6,7	8,3	16,4	17,3
	CO	5,6	0,3	4,9	4,4	6,0	14,3	15,1
40–49	MAE	9,0	8,5	5,9	6,1	12,7	13,6	14,9
	CO	–1,5	–6,1	–2,4	–4,1	5,4	12,6	11,6
50–59	MAE	13,6	17,2	11,5	14,2	12,2	6,7	8,3
	CO	–9,8	–17,2	–11,5	–14,4	–4,2	3,1	3,9
60–70	MAE	21,4	26,2	19,2	23,4	11,6	5,5	6,0
	CO	–21,4	–25,7	–19,2	–23,4	–7,3	–5,2	–4,7

Примечания. Об ограничениях проведенных сопоставлений см. текст в разделе «О перспективах повышения точности и сопоставимости оценок возраста».

Notes. See text in section «On prospects for improving the accuracy and comparability of age estimates» for limitations of comparisons».



Рисунок 2. Возрастная структура исследованной серии
Figure 2. Age-at-death structure of the series

Независимо от выбранного метода во всех случаях наблюдается одна и та же тенденция – завышение возраста в младшей возрастной группе и занижение в старшей. В некоторых работах эта проблема обозначается как «тяготение к середине» («*attraction to the middle*», «*middle-aged spread*») [Falys, Lewis, 2010], однако вполне очевидно, что у наблюдаемого явления есть другой общеупотребимый термин – регрессия к среднему. Это один из наиболее известных статистических феноменов, впервые описанный Ф. Гальтоном. Фактически он проявляется во всех случаях, в которых взаимосвязь между интересующими исследователя переменными является неидеальной, и при этом одни переменные используются для прогнозирования значений других. Максимальные и минимальные баллы облитерации швов наблюдаются не у самых пожилых и самых молодых индивидов в выборке, а у тех, возраст которых несколько смещен к среднему значению. Иными словами, результат обусловлен неполной положительной корреляцией между календарным (паспортным) и биологическим возрастом. Проблема усугубляется тем, что в течение жизни человека биологический возраст имеет тенденцию все больше отклоняться от календарного.² Эффект проявляется и в изменчивости возрастных признаков черепа, о чем вполне наглядно свидетельствуют данные, приведенные в таблице 3. Отдельным

² Влияние этого фактора, впрочем, зависит во многом от принятого метода оценки: в некоторых случаях применение априорных модельных распределений, напротив, позволяет вновь давать более узкие доверительные интервалы при оценке возраста индивидов, преодолевших порог в 60–70 лет [Milner, Boldsen, 2012].

источником таких расхождений может быть также селективный отбор – фактор, который мог бы объяснить присутствие в некоторых паспортизированных сериях значительного числа пожилых людей с открытыми швами [Key et al., 1994]. Проблема этого объяснения заключается в его умозрительном характере, поскольку сами по себе вероятные преимущества позднего заращения швов совершенно не очевидны.

Влияние ошибки в оценке возраста на возрастную структуру выборки

Оценки рассчитанного по задокументированным данным и установленным оценкам среднего возраста индивидов в выборке совпадают. Он составляет 41–43 года по оценкам авторов (в зависимости от определения условной верхней границы для открытых интервалов) и 42 года по задокументированным данным.

Возрастная структура выборки, основанная на индивидуальных значениях, рассчитанных двумя авторами, представлена на рисунке 2. Для наглядности доли возрастных групп по информации о прогнозируемых оценках в интервалах 60–64, 65–69, 70–74 лет рассчитаны условно – для открытых интервалов верхняя граница была принята равной 70 годам (возраст самого старшего индивида в выборке). Хорошо заметно завышение первым автором долей индивидов 20–24 лет, а также старше 55 лет. По оценкам второго автора, напротив, индивиды моложе 20 лет в выборке отсутствуют, доля индивидов старше 55 лет ниже реальной, а число индивидов зрелого возраста заметно превышает ожидаемую величину.

Возрастная структура, полученная путем усреднения оценок двух исследователей наиболее близка к реальной. Если средняя абсолютная разница между реальной и установленной долями каждой из пятилетних возрастных групп составляет 3,0 и 2,7% для авторов 1 и 2, то для усредненной структуры она составляет всего 1,6%

В целом наблюдаемые отклонения могут показаться незначительными, однако в некоторой степени это впечатление может оказаться обманчивым из-за особенностей реальной возрастной структуры выборки: распределение имеет унимодальную форму (если не считать небольшого и статистически незначимого пика в группе 65–69 лет), при этом мода приходится на интервал 35–39 лет.

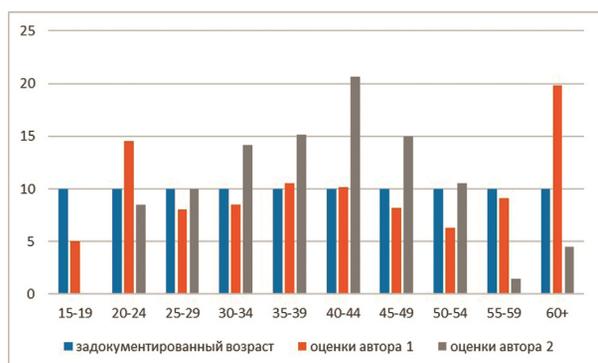


Рисунок 3. Возрастная структура модельной выборки с равным представлением индивидов в каждом пятилетнем интервале

Figure 3. Age-at-death structure of the model sample with equal representation of individuals in each five-year interval

Для проверки этого предположения из имеющейся выборки была сформирована модельная выборка, в которой все возрастные пятилетние интервалы представлены равным числом индивидов. Чтобы ограничить влияние открытых интервалов в оценках авторов на результаты, индивиды старше 60 лет составили финальную когорту. Модельная выборка была создана путем случайного отбора в каждый пятилетний интервал 10 индивидов с соответствующим задokumentированным возрастом. Выборка формировалась с возвращением: поскольку некоторые возрастные группы представлены менее чем 10 индивидами, у каждого индивида сохранялся шанс повторно войти в выборку после уже состоявшегося отбора.

При анализе модельной выборки проявилось то же влияние систематической ошибки в оценках авторов, которое наблюдалось в исходном варианте, но отклонения оказались более выразительными (рис. 3). В возрастных распределениях, построенных на основании оценок авторов, наблюдаются выразительные «пики смертности», которые, однако, не находят соответствия в реальной возрастной структуре выборки. В публикациях, посвященных палеодемографическим исследованиям, пики такого типа часто получают содержательную интерпретацию с историко-демографических позиций, и вопрос о надежности их установления авторами, как правило, не ставится.

Общее влияние систематической ошибки в оценке возраста индивидов, относящихся к конкретным возрастным когортам, оказалось контринтуитивным. В соответствии с приведенными выше данными у обоих авторов наблюдалась

одинаковая тенденция: занижение возраста пожилых индивидов и завышение возраста молодых. Однако если форма возрастного распределения, установленная по оценкам второго автора, полностью соответствует ожидаемой в рамках эффекта регрессии к среднему, то структура, определенная по оценкам первого автора, напротив, демонстрирует завышение долей молодых и пожилых индивидов. В этом случае также можно наблюдать слабое проявление регрессии к среднему, но сложение случайных и систематических погрешностей в оценках возраста привело на первый взгляд к неожиданному результату. Причина различий заключается в том, что для того чтобы прогнозировать влияние ошибки на структуру выборки недостаточно информации о величине средних отклонений оценок от задokumentированного возраста в конкретных возрастных группах. Необходимо также учитывать отклонения задokumentированного возраста от оценок для возрастных групп, состав которых установлен по оценкам исследователей.

О перспективах повышения точности и сопоставимости оценок возраста

Точность оценок – как визуальных, так и строго формализованных – в некоторой степени зависит от возрастной структуры не только той выборки, с которой работают исследователи, но и той паспортизированной серии, на изучении которой опирается использованная авторами методика. Впервые внимание на этой проблеме заострили французские палеодемографы Ж.П. Боке-Аппель и К. Массе [Bocquet-Appel, Masset, 1982; 1996]. В зависимости точности оценок от структуры выборки заключается причина, по которой некорректно сравнивать коэффициенты корреляции, рассчитанные между задokumentированным возрастом и оценками, полученными в рамках отдельных методов и основанными на анализе разных серий. Но даже если оценки опираются на анализ серий со сходным возрастным распределением, точность первых будет различаться при работе с разными возрастными группами внутри выборки. Это очевидное наблюдение часто игнорируется исследователями при сравнительном анализе эффективности разных подходов.³

³ Например, сведенные в единую таблицу данные о средней величине коэффициента корреляции, приведенные в обзорной работе Ruengdit et al., 2020, в действительности малоинформативны для корректного сравнения методик.

И все же представляется, что качество оценок возраста, основанных на визуальном подходе, может быть несколько увеличено независимо от структуры выборки, с которой придется работать исследователю. Под качеством оценок в данном случае понимаются две разные характеристики: точность и сопоставимость. Теоретически для улучшения этих характеристик могут быть использованы несколько подходов, из которых авторы для предварительной оценки выделили три: 1) повторное изучение серии; 2) увеличение возрастных интервалов (отказ от 10- и тем более 5-летних интервалов); 3) фиксация признаков в балловых шкалах. Первый подход может позволить увеличить точность оценок, тогда как два остальных – сопоставимость оценок разных исследователей.

1. Существует множество исследований, результаты которых свидетельствуют о том, что усреднение оценок, вынесенных разными экспертами, позволяет точнее оценить исследуемый параметр, чем оценки отдельных людей [Hertwig, 2012]. Этот подход получил название «мудрость толпы», а его эффективность неоднократно получала подтверждение в задачах определения некоторых количественных параметров при условии, что участники не испытывают внешнего влияния при вынесении решения, и ошибки носят случайный характер. Увеличить точность можно также прибегнув к эффекту «мудрости внутренней толпы». Эффект может быть достигнут путем усреднения оценок, установленным одним и тем же исследователем, если они были вынесены с большим временным интервалом [Vul, Pashler, 2008] или если при повторной оценке исследователь мысленно принял позицию человека, не согласного с обоснованием его исходной оценки [Van de Calseyde, Efendic, 2022]. В первом случае предполагается, что человек не имеет возможности отталкиваться от первоначальных оценок, тогда во втором они являются важным элементом процесса решения задачи.

О пользе повторного определения возраста и усреднения полученных оценок свидетельствуют результаты небольшого теста. Девять лет назад первый автор участвовал в семинаре, участники которого оценивали пол и возраст 15 черепов русских с задокументированным полом и возрастом, происходящих из той же коллекции К.З. Яцуты (№ 1830). Средняя величина MAE для оценок автора составила 6.9 года по определениям 2015 года и 7,0 по определениям

2024 года. Усреднение оценок разных лет привело к снижению погрешности до 5,7 года. Среднее отклонение в оценках возраста среди участников семинара 2015 года составило 8,4 года. Однако между общими усредненными оценками возраста и задокументированным возрастом MAE составило 6,6 года – меньше, чем для оценок любого отдельно взятого участника. Таким образом, положительный эффект «мудрости толпы» предварительно подтверждается и при оценке возраста по черепу, хотя возможности применения подхода требуют более подробного изучения.

2. Наши субъективные представления о возможностях определения возраста в узких интервалах очевидным образом не соответствуют действительности. Эта проблема не является специфической для антропологии. Склонность к излишней самоуверенности в оценках параметров и соответствующих доверительных интервалов свойственна людям независимо от сферы их занятости [Soll, Klayman, 2004; McKenzie et al., 2008]. Увеличение интервалов при оценке возраста несомненно увеличит как вероятность попадания в них реального значения, так и сопоставимость заключений разных исследователей. Меньше всего ошибок нас ожидает при использовании интервалов «ребенок», «подросток», «взрослый», и в рамках некоторых подходов такая разбивка, например, достаточна для проведения палеодемографического анализа [Bocquet-Appel, Masset, 1996], однако вряд ли большинство исследователей когда-либо согласится ограничиваться такими определениями⁴. Симметричное расширение интервалов, примененное к опубликованным данным, может оказаться эффективным при анализе палеодемографических данных [см., например: Ширококов, 2020]. Однако такое расширение само по себе не позволяет решить проблему систематической ошибки в определении возраста. Вероятно, расширение интервалов должно быть несимметричным и для внесения поправок могут использоваться оценки в рамках первого подхода или некоторая информация об априорном распределении возрастных групп.

⁴ Кроме того, на самом деле в антропологии существует проблема сопоставимости оценок даже для такой категории как «взрослый», в т.ч. на уровне определения базового набора дифференцирующих биологических характеристик скелета [Falys, Lewis, 2010]

3. Фиксация признаков в некоторой стандартизованной шкале, возможно, напрямую не повлияет непосредственно на точность оценок возраста, однако она позволит увеличить сопоставимость заключений разных авторов. Как было показано выше, оценки признаков обнаруживают несколько большую межисследовательскую согласованность по сравнению с оценками возраста. Кроме того, использование балловых оценок позволяет развивать новые подходы в палеодемографии, которые опираются на некоторые предположения об априорном распределении и позволяют оценивать возрастную структуру выборки целиком, минуя этап индивидуальных оценок возраста. Наконец, фиксация в баллах степени развития некоторого набора признаков, имеющих доказанную возрастную изменчивость, позволяет более объективно подходить к вопросу о возрастной изменчивости дискретно-варирующих или метрических признаков, которые антропологи используют при межгрупповых сопоставлениях.

Заключение

Поскольку точность субъективно-визуального подхода не уступает классическим методам оценки возраста по степени облитерации швов, преимущества первого могут показаться вполне очевидными. Формализованный подход выглядит негибким, потому что никогда не учитывает всех признаков возрастных изменений черепа, доступных глазу: легко представить случай, в котором согласно схеме облитерации швов индивид формально должен быть отнесен к группе *adultus*, хотя дегенеративные изменения костей свидетельствуют о том, что череп принадлежит индивиду зрелого или старческого возраста.

Кажется, что решение этой проблемы очевидно – при оценке возраста следует пользоваться комплексными формализованными методами, учитывающими возрастную изменчивость разных элементов скелета [Lovejoy et al., 1985; Boldsen et al., 2002; Milner, Boldsen, 2012]. И эта стратегия действительно сегодня используется большинством исследователей (но, по всей видимости, без строгой фиксации признаков). Если в распоряжении исследователя находится только череп, теоретически он также может применить более гибкий подход к оценке возраста, даже если при этом вынужден опираться преимущественно на анализ степени облитерации швов. На первом этапе оценивается принадлеж-

ность индивида к большой возрастной группе (старше или моложе 50 лет), после чего проводится оценка набора признаков, наиболее подходящего для более узкой интервальной оценки возраста [Perizonius, 1984; Key et al., 1994].

Однако на проблему можно взглянуть и с другой стороны. В действительности, к какому бы методу исследователи не прибегали, процедура оценки возраста никогда не сводится к одному этапу и не является строго формализованной. Работая с любой выборкой, мы разделяем скелеты детей и взрослых, и уже затем выбираем наиболее оптимальные методы для проведения анализа. Мы с осторожностью устанавливаем оценки скелетов, отличающихся некоторыми явными патологическими особенностями. Мы определяем влияние степени сохранности останков на возможности определения характеристики и оцениваем целесообразность определения признаков в спорных случаях. Субъективно-визуальный подход является не альтернативой для строгой формализованной оценки, а его важным дополнением. Несомненно, стандартизация методов позволяет увеличить сопоставимость оценок разных авторов, а также корректнее подходить к решению отдельных популяционно-антропологических и палеодемографических задач. Несомненно, отбор оптимальных методик, доказавших свою эффективность по результатам независимых исследований (и значит, невозможный без формализации), позволяет увеличить точность оценок возраста. Но несомненно и то, что качество оценок, выдвигаемых в рамках субъективно-визуального подхода, также может быть увеличено, а поиск тех способов организации процедуры исследований, которые могут быть использованы для решения этой проблемы, представляется авторам самостоятельной нетривиальной задачей.

Благодарности

Публикация подготовлена при финансовой поддержке гранта РНФ №24-28-01050 «В поисках эффективных методов оценки пола и возраста по черепу».

Библиография

Герасимов М.М. Восстановление лица по черепу (Современный и ископаемый человек). М.: Изд-во АН СССР. 1955. 586 с.

Звягин В.Н. Преждевременное зарастание черепных швов в судебно-медицинском аспекте // Судебно-медицинская экспертиза, 1976. №3. С. 18–23.

Никитюк Б.Н. О закономерностях облитерации швов на наружной поверхности мозгового отдела черепа человека // Вопросы антропологии, 1960. №2. С. 115–121.

Широбокое И.Г. Влияние миграций на палеодемографическую характеристику населения России XVII–XIX вв. // Сибирские исторические исследования, 2020. № 4. С. 157–181. DOI: 10.17223/2312461X/30/8.

Информация об авторах

Широбокое Иван Григорьевич, к.и.н.; ORCID ID: 0000-0002-3555-7509; ivansmith@bk.ru;

Пугачева Екатерина Владимировна, ORCID ID: 0000-0003-3948-2759; pugacheva.kari@yandex.ru.

Поступила в редакцию 10.05.2024,
принята к публикации 03.06.2024

Shirobokov I.G.¹⁾, Pugacheva E.V.²⁾

¹⁾ Peter the Great Museum of Anthropology and Ethnography (Kunstkamera) RAS, Department of Physical Anthropology, Universitetskaya emb., 3, Saint Petersburg, 199034, Russia

²⁾ L'École normale supérieure, 45, rue d'Ulm / 29 rue d'Ulm / 24 rue Lhomond, F-75230 Paris cedex 05, France

ON THE ACCURACY OF VISUAL AGE ESTIMATION FROM THE ADULT SKULL (METHODOLOGICAL ASPECTS)

Introduction. There are several standardized methods for estimating the age of a skull. Most of these methods are based on the analysis of suture obliteration and the tooth wear scoring. However, many anthropologists prefer a more subjective approach, relying on general impressions without using a set of standardized criteria. This study aimed to assess the effectiveness of a visual method for age estimation and reconstruction of age-at-death structure in a skeletal sample.

Materials and methods. The study was based on a series of 116 skulls from the early 20th century collected by the Peter the Great Museum of Anthropology and Ethnography (Kunstkamera). These specimens had documented sex and age information. Two researchers independently assessed the age of the skull specimens and recorded the degree of suture fusion on the cranial vault as well as the level of tooth wear on the occlusal surfaces. The correlation between age and estimated scores was calculated using Spearman's rank correlation coefficient. The discrepancy between estimated and actual ages was measured by calculating the mean absolute error (MAE) and systematic error (SE) as the average difference between documented and estimated ages for the entire sample as well as for each age group. Intraclass correlation coefficients were used to assess the consistency of the authors' estimates.

Results. The authors' estimates showed moderately high agreement among themselves and a moderate positive correlation with actual age. The accuracy of the visual assessments was found to be comparable with that of more formalized methods for assessing the degree of suture obliteration. The estimates also exhibited the phenomenon of regression to the mean, with individuals in younger cohorts being systematically overestimated in terms of age and those in older cohorts being underestimated. The accuracy of determining the age-at-death distribution depends to some extent on the actual characteristics of the sample structure. Averaging estimates from different authors or several estimates from the same author, repeated over a large time interval, makes it possible to bring estimates closer to real data.

Conclusion. Increasing interobserver agreement of age estimates can be achieved by fixing traits on the same point scales, as well as by increasing age intervals. The accuracy of estimates can be improved by repeated age estimation as well as the “wisdom of the crowd” effect.

Keywords: cranial vault sutures; age estimation; age-at-death structure; regression to the mean; the wisdom of the crowd

DOI: 10.55959/MSU2074-8132-24-4-8

References

- Gerasimov M.M. *Vosstanovlenie litsa po cherepu (Sovremenniy i iskopaemyi chelovek)* [Facial reconstruction from a skull (Modern and fossil humans)]. Moscow: AS USSR Publ., 1955. 586 p. (In Russ).
- Zvyagin V.N. Prezhdevremennoe zarastanie cherepnykh shvov v sudebno-meditsinskom aspekte [Premature suture closure of the skull in forensic-medical aspect]. *Sudebno-Meditsinskaya Ekspertiza* [Forensic Medical Expertise], 1976, 3, pp. 18–23. (In Russ).
- Nikitjuk B.N. O zakonomernostyakh oblitieratsii shvov na naruzhnoy poverkhnosti mozgovogo otdela cherepa cheloveka [On the patterns of suture obliteration on the outer surface of the human cranial part]. *Voprosy Antropologii* [Questions of Anthropology], 1960, 2, pp. 115–121. (In Russ).
- Shirobokov I.G. Vliyanie migratsiy na paleodemograficheskuyu kharakteristiku naseleniya Rossii XVII-XIX vv. [Impact of migrations on paleodemographic characteristics of the Russian population, 17th to 19th centuries]. *Sibirskie istoricheskie issledovaniya* [Siberian Historical Research], 2020, 4, pp. 157–181. (In Russ).
- Acsadi G., Nemeskeri J. *History of human life span and mortality*. Budapest, 1970. 196 p.
- Bocquet-Appel J.P., Masset C. Farewell to paleodemography. *J. Hum. Evol.*, 1982, 11, pp. 321–333.
- Bocquet-Appel J.P., Masset C. Paleodemography: expectancy and false hope. *Am. J. Phys. Anthropol.*, 1996, 99, pp. 571–583. DOI: 10.1002/(SICI)1096-8644(199604)99:4<571::AID-AJPA4>3.0.CO;2-X.
- Boldsen J.L., Milner G.R., Konigsberg L.W., Wood J.W. Transition analysis: a new method for estimating age from skeletons. Hoppa RD, Vaupel JW, editors. *Paleodemography: age distributions from skeletal samples*. Cambridge, Cambridge University Press, 2002, pp. 73–106.
- Falys C.G., Lewis M.E. Proposing a way forward: A review of standardisation in the use of age categories and ageing techniques in osteological analysis (2004–2009). *Int. J. Osteoarchaeology*, 2010, 21 (6), pp. 704–716. DOI: 10.1002/oa.1179.
- Galera V., Ubelaker D.H., Hayek L. Comparison of macroscopic cranial methods of age estimation applies to skeletons from the Terry collection. *Journal of Forensic Sciences*, 1998, 43, pp. 933–939.
- Hertwig R. Tapping into the wisdom of the crowd – with confidence. *Science*, 2012, 336(6079), pp. 303–304. DOI:10.1126/science.1221403.
- Key C.A., Aiello L., Molleson T. Cranial suture closure and its implications for age estimation. *Int. J. Osteoarchaeology*, 1994, 4, pp. 193–207.
- Lovejoy C.O., Meindl R.S., Mensforth R.P., Barton T.J. Multifactorial determination of skeletal age at death: a method and blind tests of its accuracy. *Am. J. Phys. Anthropol.*, 1985, 68 (1), pp. 1–14.
- McKenzie C.R.M., Liersch M.J., Yaniv I. Overconfidence in interval estimates: What does expertise buy you? *Organizational Behavior and Human Decision Processes*, 2008, 107 (2), pp. 179–191. DOI: 10.1016/j.obhdp.2008.02.007.
- Meindl R.S., Lovejoy C.O. Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *Am. J. Phys. Anthropol.*, 1985, 68, pp. 57–66.
- Milner G.R., Boldsen J.L. Transition analysis: a validation study with known-age modern American skeletons // *Am. J. Phys. Anthropol.*, 2012, 148, pp. 98–110.
- Perizonius W.R.K. Closing and non-closing sutures in 256 crania of known age and sex from Amsterdam (a.d. 1883–1909). *J. Hum. Evol.*, 1984, 13 (2), pp. 201–216.
- Ruengdit S., Prasitwattanaseree S., Mekjaidee K., Sinthubua A., Mahakkanukrauh P. Age estimation approaches using cranial suture closure: A validation study on a Thai population. *J. Forensic and Legal Medicine*, 2018, 53, pp. 79–86. DOI: 10.1016/j.jflm.2017.11.009.
- Ruengdit S., Case D. T., Mahakkanukrauh P. Cranial suture closure as an age indicator: A review. *Forensic science international*, 2020, T. 307, pp. 110–111. DOI: 10.1016/j.forsciint.2019.11011.
- Shrout P.E., Fleiss J.L. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 1979, 86 (2), pp. 420–428.
- Soll J.B., Klayman J. Overconfidence in Interval Estimates. *J. Experimental Psychology: Learning, Memory, and Cognition*, 2004, 30 (2), pp. 299–314. DOI: 10.1037/0278-7393.30.2.299.
- Todd T.W., Lyon D. Endocranial suture closure. Its progress and age relationship. Part I. – adult males of white stock. *Am. J. Phys. Anthropol.*, 1924, 7 (3), pp. 325–384.
- Van de Calseyde P.P., Efendić E. Taking a disagreeing perspective improves the accuracy of people’s quantitative estimates. *Psychological Science*, 2022, 33, pp. 971–983. DOI: 10.1177/09567976211061321.
- Vul E., Pashler H. Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science*, 2008, 19 (7), pp. 645–647. DOI: 10.1111/j.1467-9280.2008.02136.x.
- Wolff K., Vas Z., Sótonyi P., Magyar L.G. Skeletal age estimation in Hungarian population of known age and sex. *Forensic Science International*, 2012, 223 (1–3), pp. 374.e1–374.e8. DOI: 10.1016/j.forsciint.2012.08.033.
- Wolff K., Hadadi E., Vas Z. A novel multidisciplinary approach toward a better understanding of cranial suture closure: the first evidence of genetic effects in adulthood. *Am. J. Hum. Biol.*, 2013, 25 (6), pp. 835–43. DOI: 10.1002/ajhb.22459.

Information about the authors

Shirobokov Ivan G., PhD.; ORCID ID: 0000-0002-3555-7509; ivansmith@bk.ru;

Pugacheva Ekaterina V., ORCID ID: 0000-0003-3948-2759; puga-cheva.kari@yandex.ru

© 2024. This work is licensed under a CC BY 4.0 license