

## A JURY OF *HER* PEERS: THE IMPACT OF THE FIRST FEMALE JURORS ON CRIMINAL CONVICTIONS\*

*Shamena Anwar, Patrick Bayer and Randi Hjalmarsson*

This article uses an original data set of more than 3,000 cases from 1918 to 1926 in the Central Criminal Courts of London to study the effect of the Sex Disqualification (Removal) Act of 1919. Implemented in 1921, this Act made women eligible to serve on English juries. Results based on a pre-post research design imply that the inclusion of women had little effect on overall conviction rates but significantly impacted conviction rates on particularly female salient cases: sex offences, violent offences with female *versus* male victims and female defendants charged with ‘other’ (largely abortion related) offences.

Measured against the long span of human history, the inclusion of women in civic and political life has been remarkably recent and remains far from complete. In the US, women have had the right to vote for fewer than 100 years and female representation in Congress has never topped 20%. A woman has not yet served as US President. Beyond the elementary issues of fairness and representation, proponents of greater female participation in government and politics argue that such inclusion alters the broad focus of policy, the collective functioning of legislative bodies, and executive decision-making.<sup>1</sup>

How the inclusion of women affects outcomes in any aspect of civic and political life is a stubbornly difficult question to answer well. The main challenge is that the selection of individuals for most positions – elected officials or political appointments – is far from random. This makes it difficult to separate the causal effect of female officials from the circumstances that lead to their appointments. That the majority of female Congress members of the US Congress represent progressive/liberal districts is an obvious reflection of this kind of selection across jurisdictions. Perhaps more subtly, this same selection problem also hampers the identification of gender differences using variation within jurisdictions. In this case, the positions of those appointed (regardless of gender) must be broadly compatible with the person or constituency doing the selecting. This tends to diminish gender differences among those selected relative to the larger population of qualified candidates. Strikingly, even the apparent gold standard of the random assignment of court cases to judges does not avoid this fundamental selection problem,

\* Corresponding author: Randi Hjalmarsson, Department of Economics, University of Gothenburg, PO Box 640, SE 40530, Gothenburg, Sweden. Email: [randi.hjalmarsson@economics.gu.se](mailto:randi.hjalmarsson@economics.gu.se).

This article would not have been possible without the tremendous efforts of our Research Assistant Michael Bekele in transcribing the original court records and without the financial support of Foundation for Economic Research in West Sweden, and Vetenskapsrådet, The Swedish Research Council, Grants for Distinguished Young Researchers.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

<sup>1</sup> A large body of correlational evidence suggests that male and female politicians differ systematically in basic policy positions, active policy focus, parliamentary approach and legislative effectiveness, implying that gender composition affects a range of substantive outcomes ([Anzia and Berry, 2011](#); [Pearson and Dancy, 2011](#); [Volden et al., 2013](#); [Wittmer and Bouche, 2013](#)).

as both female and male judges must have views and judicial approaches that allowed them to be appointed in the same district in the first place.<sup>2</sup>

In countries that follow the tradition of English common law, juries for criminal and civil trials constitute perhaps the most broad-based deliberative bodies in formal civic life. The right to a trial by a jury of one's peers is one of three laws (among 63 original clauses) to survive more or less intact over the eight centuries since the English *Magna Carta* was first written in 1215. Of course, the definition of 'peers' has changed tremendously over time, as eligibility criteria for jury service have become increasingly inclusive along the dimensions of age, wealth, race and gender. Conditional on eligibility, however, citizens are often drawn at random for jury duty, which provides a basis for addressing the fundamental selection issue described above.

In this article, we study one of the most substantial changes in eligibility criteria – the introduction of female jurors. Women became eligible to serve on English juries with the passing of the Sex Disqualification (Removal) Act in December of 1919, though they were not officially added to the jury pool until 1921. At the same time, similar movements were taking off around the US; Utah became the first state in 1898 to authorise women to participate on juries but three states (Alabama, Mississippi and South Carolina) banned female jurors as late as 1962.

We examine the impact of the first female jurors on convictions in criminal cases in London using an original data set that we have created from hand-written court records of the First and Second Courts of the Old Bailey (the Central Criminal Court for the city of London and surrounding county of Middlesex) from 1918 to 1926. We do not collect records for earlier years given the differential functioning of the courts during World War I. From these records, we have extracted detailed information about the case (defendant names, charges, victim names, pleas or jury trials, verdicts and sentences), as well as the names of all individuals seated on the jury. The final data set consists of more than 3,000 observations.

Our focus on the historical period surrounding the introduction of the first female jurors in England is motivated by several important (related) considerations.<sup>3</sup> First, even though eligible citizens may be randomly called for jury duty, the pre-trial jury selection process (*voir dire*) provides the attorneys and judge considerable discretion over which members of the jury pool are actually seated for the trial, reintroducing a version of the fundamental selection issue described above. Second, variation in the composition of modern juries is typically centred at a point of approximate gender balance. Thus, even if research designs exploit random variation in the composition of the jury pool to deal with pre-trial selection, the results are generally only informative about variation near this point of gender balance – e.g. identifying the difference in outcomes

<sup>2</sup> The findings concerning female judges in the criminal courts are quite mixed. Even in the same jurisdiction, contradictory results have been found; Johnson (2014) finds female judges in Pennsylvania are more lenient than male judges while Steffensmeier and Herbert (1999) find that they impose harsher sentences. More generally, George (2001) concludes that the empirical literature has failed to find any broad and consistent gender-based differences in judging. In fact, many researchers have concluded that judge characteristics, of which gender is just one example, simply do not explain much of the variation in judicial decisions (Ashenfelter *et al.*, 1995), which is entirely consistent with our characterisation of the fundamental selection problem above.

<sup>3</sup> We study the introduction of female jurors in England (rather than the US, e.g.) because the data that we have gathered provide direct measures of the impact of the reform on the extent to which females were actually seated on juries. Comparably rich US data are not available.

for 12-person juries with six *versus* seven female members. In contrast, our analysis estimates the impact of including female jurors on what had been all-male juries, providing a sharper and more meaningful comparison for assessing the broad impact of female representation.

The question of whether female representation on juries affects verdicts has been raised by the popular press both today and at the time of the ‘great experiment’ – the term used by newspapers during the period to characterise the new eligibility of female jurors. For instance, a jury of six women decided George Zimmerman was not guilty in the shooting of an unarmed black teenager, Trayvon Martin, in 2014; subsequently, almost every media outlet in the country included a headline questioning whether the extreme gender composition of the jury affected the verdict. During the peak of the women’s jury movement in the United States, Susan Glaskell wrote a short story titled ‘*A Jury of Her Peers*’ that depicts how men and women perceive things differently in the household; though the story does not take place in a courtroom, it depicts how male and female jurors might evaluate evidence through different lenses, particularly in the case of a female defendant.<sup>4</sup>

The existing empirical literature has used experiments (mock juries) and post-trial juror surveys to study whether female jurors have an impact on trial outcomes.<sup>5</sup> The results of the mock jury studies are both inconsistent and often limited in their generalisability given the experimental simplifications, non-representative subjects, and lower stakes compared to the real world.<sup>6</sup> Studies using post-verdict survey data in the US generally find little or no evidence of a gender effect.<sup>7</sup> Few studies of UK juries exist given the 1981 Contempt of Court Act, which made it illegal to survey jurors. Baldwin and McConville (1979) conducted one of the first (and only) studies of English juries using a sample of 500 trials in Birmingham, finding no evidence that jury composition, including gender, affects verdicts. The Ministry of Justice has recently begun to address this knowledge gap concerning the impartiality of English juries, particularly with respect to race (Thomas and Balmer, 2007). Conducting mock trials with actual jurors at the Crown Courts, Thomas (2010) finds female jurors presented with a male defendant charged with bodily harming a male victim were significantly more likely to vote to convict at the start of deliberations, but also more likely to change their views during deliberations – such that there was no significant difference between male and female

<sup>4</sup> First published in 1917, ‘*A Jury of Her Peers*’ takes place at the Iowa farm of Minnie Wright, who is charged with murdering her husband. The sheriff, prosecutor and a neighbour (all men) come to gather evidence to support the charge and two women (the wives of the sheriff and neighbour) accompany them. While the men find nothing unusual in the kitchen, the women note a disordered household, an irregular quilting pattern, and find a strangled canary. These clues lead them to discuss Minnie’s abusive and authoritarian husband, and to sympathise with her desperation. In essence, these women decide she is not guilty and not to reveal the motive to the investigators.

<sup>5</sup> See Devine *et al.* (2000) for a brief review of this literature.

<sup>6</sup> Mock jury studies exist that find females are:

- (i) pro-conviction (Hastie *et al.*, 1983; Hegelson and Shaver, 1990; Gabora *et al.*, 1993; Fischer, 1997);
- (ii) pro-acquittal (Fitzgerald and Ellsworth, 1984; Thompson *et al.*, 1984); and
- (iii) no different (Simon, 1967; Cowan *et al.*, 1984; LaFree *et al.*, 1985; Goodman *et al.*, 1998).

<sup>7</sup> Studies finding no effect of seated female jurors include Bridgeman and Marlowe (1979), Moran and Comfort (1982), Garvey *et al.* (2004), and most recently, Lehmann and Blair Smith (2013).

conviction rates in the final vote. Mills and Bohannon (1980) report similar findings for the US.

Our primary empirical analysis is based on a comparison of verdicts from the full set of trials before and after the introduction of the reform. As highlighted above, this reduced form approach focuses on the margin of the basic inclusion (or not) of female jurors and avoids the potential problem of pre-trial jury selection by not explicitly conditioning on the number of seated female jurors in the post-reform period.<sup>8</sup> The estimated effect encompasses many potential mechanisms, the most obvious of which is that eligibility leads directly to female representation on juries, thereby affecting deliberations and decisions. But the reform may also indirectly affect verdicts even in cases in which females are not seated on the jury. This could occur, for instance, if the judge makes comments to the entire jury pool about whether the nature of the case is 'sensitive' for female ears or if pre-trial selection is focused on striking potential female jurors, which could affect the resulting distribution of seated male jurors.

While the Sex Disqualification Act was passed in 1919, it was not implemented for jurors until 1921 because it took a year to write the rules governing the inclusion of female jurors and to update the jury books from which the pools were drawn. We find that more than half of seated juries in 1921 had at least one female juror and that more than 80% of cases in the following years had female jurors; there were no females seated prior to 1921. In addition, though males clearly still dominated the jury, more than 20% of the post-reform juries had at least three female jurors (25% of seated jurors). Our analysis of seated juries also indicates that in the post-reform period, there was some 'selection' of females according to case characteristics, especially for sex offences. This kind of differential pre-trial selection raises the kinds of concerns that motivate our emphasis on pre-post reform comparisons in the first place.

Because our main results are based on comparing outcomes before and after the reform, it is important to assess whether other changes may have occurred regarding crime or the criminal justice system at the time of the reform. While we have not identified any concurrent policy changes,<sup>9</sup> there does appear to be a small change in the distribution of crimes for which defendants were charged following the reform, at least in the Second Court. As a result, we condition our entire analysis on a large and detailed set of controls for criminal charges and other case characteristics. A related possibility is that changes in plea behaviour or even prosecutor charge behaviour altered the set of cases that reached a jury trial following the reform. Conditional on the set of criminal charges, however, there is no evidence of any change in plea behaviour coinciding with the reform. In addition, we show that there is no change in sentence length for either plea cases or jury trials after the reform, suggesting that there is no change in the observable or unobservable determinants of sentence length.

Our analysis yields a number of key findings, which are robust to the inclusion of a full set of controls for crime and case characteristics, the consideration of alternative

<sup>8</sup> In contrast, this margin is easier to study in modern day data in the context of race, for instance, given that many juries and jury pools do not include any black members (Anwar *et al.*, 2012).

<sup>9</sup> The Representation of the People Act of 1918 made females over age 30 eligible to vote. This right was extended to females between ages 21 and 30 by the Representation of the People (Equal Franchise) Act of 1928.

post-reform windows, and alternative dependent variables. First, bundling all cases together, the reform did not have a significant impact on the likelihood of conviction. This overall result, however, masks several countervailing findings for specific crime categories. In particular, female representation on juries significantly increased conviction rates for sex offence cases by 16 percentage points and decreased those for property and violent crimes by 10 and 13 percentage points, respectively. The magnitude of these effects is substantial, especially in the light of the fact that males continued to constitute a large majority (more than 80%) of jurors post-reform. Interestingly, the decrease in conviction rates for violent crimes as a whole also conceals opposing results based on the gender of the victim. In fact, prior to the reform the conviction rate differential between male and female victim violent crime cases was essentially zero; post-reform this conviction differential between female and male victim cases increased to 20 percentage points. Finally, we find a reduction in convictions for female defendants charged with ‘other’ offences after the reform, which is quite large and sometimes significant; notably, at least one-third of the female defendants in this category are charged with an abortion related offence.

We complement these pre-post design results with a secondary analysis that exploits the fact that seated juries were frequently used for multiple trials. In particular, we analyse the impact of seated female jurors for the subset of cases in which the jury was carried over from a previous trial. We expect issues involving the non-random seating of female jurors to be muted in these cases and, in fact, the gender composition of repeated juries’ is not correlated with the gender of the victim in violent crime cases. This analysis reveals a pattern of results that is remarkably consistent with the initial pre-post reform analysis, indicating that adding females to the seated jury sharply increased conviction rates for violent crime cases against women *versus* men.<sup>10</sup>

Taken as a whole, the results of our analysis imply that female representation on juries substantially affects the likelihood of conviction for a subset of cases – sexual and violent crimes – in which female jurors might have viewed the alleged behaviour or its impact on the victim from a different perspective than their male counterparts. While this study is based on data from nearly a century ago, the magnitude of the results raises concerns about the basic fairness and effectiveness of modern criminal justice systems that exclude (or severely limit the role of) women. Such concerns are not limited to societies that completely exclude women from serving as judges and jurors but extend to courts or institutions, such as the military, where the historical overrepresentation of men in positions of authority continues to limit the role of women in judicial decision-making.<sup>11</sup>

<sup>10</sup> These results are broadly consistent with results reported for Sweden in [Anwar et al. \(2015\)](#), which finds that lay judges (*nämndemän*) from the feminist Vänster party are significantly more likely to convict defendants when the victim of violent crime is female. Criminal trials in Sweden are decided by a judge and three *nämndemän*, who are nominated by political parties in proportion to the party’s performance in the most recent election.

<sup>11</sup> The male-dominated command structure of the US military, for example, has long been viewed as a major factor in the extraordinarily low rate of successful prosecutions of sexual assaults. The [Department of Defense \(2014\)](#) estimates that approximately 5,000 of 25,000 incidents of sexual assaults were reported in 2013 and, of these, only 375 (less than 2% of all sexual assaults and 8% of those reported) were successfully prosecuted.

The remainder of the article proceeds as follows. [Section 1](#) provides institutional details regarding the English jury, and the context of the female jury reforms. [Section 2](#) describes the data and [Section 3](#) presents an analysis of the first stage impact of the reform on the gender of seated jurors and the selection of jurors. [Section 4](#) presents the basic pre-post design and findings, as well as robustness tests and an analysis of whether seated jurors are driving the results. [Section 5](#) concludes.

## 1. Institutional Background

### 1.1. *The English Jury*

The origins of the English jury date back to the *Magna Carta* in 1215, which provided the first set of rules to apply throughout the land. It states ‘No free man shall be seized or imprisoned or stripped of his rights or possessions, or outlawed or exiled, or deprived of his standing in any other way ... except by the lawful judgement of his equals or by the law of the land’. Clearly, the implementation of this law has changed significantly with the passage of time.<sup>12</sup>

The cases analysed in this study are drawn from London and the surrounding suburbs in the 1920s. The rules in place during the 1920s differ significantly from those in place today due, in particular, to the Juries Act of 1974, which sets out the rules governing the modern English jury system. Most notably, the 1974 Act changed the eligibility requirements to serve as a juror from ones that were based on wealth to being a registered voter.

The following is a brief overview of the jury system during our study period, which was based on the Juries Act of 1825. To qualify for jury service one had to be: male, between the age of 21 and 60, residing in England, and a holder of land or wealth of various thresholds.<sup>13</sup> Explicitly disqualified were foreign aliens and justices of the peace, and Westminster inhabitants could not be jurors for cases from Middlesex (the suburbs of London).

A master list of qualified men was to be created annually in each parish by the churchwardens and overseers of the parish. The list indicated each individual’s name as well as some additional information (e.g. title, business, and nature of qualification) and was to be posted on church doors for the first three Sundays of September in each year to give parish members the chance to object. The final list was kept by the clerk of peace and copied into a book to be delivered to the sheriff, titled ‘The jurors book for the year XXXX’. This master list was brought into use on 1 January of the following year, and used for one year.<sup>14</sup>

<sup>12</sup> See [Cockburn and Green \(1988\)](#) for a detailed history of the criminal trial jury in England dating back to the year 1200 and [Bentley \(1998\)](#) for a more focused discussion of the English criminal justice system in the nineteenth century.

<sup>13</sup> More specifically, a juror should:

- (i) possess an income of 10 pounds per year from real estate or rent charge; or
- (ii) 20 pounds per year from a leasehold of not less than 21 years; or
- (iii) should be a householder living in premises rated no less than 20 pounds per year (30 pounds in London and Middlesex); or
- (iv) should occupy a house with no fewer than 15 windows.

<sup>14</sup> Unfortunately, to the best of our knowledge, these juror books have not survived.



During the time period studied, the London courts held monthly sessions, which typically lasted for two to three weeks. Jurors received a summons 10 days before the required day of attendance. Unfortunately, little is known about how the pool of summoned jurors was selected from the jury book (Langbein, 1987). But, the Jury Act does include guidelines on how an empanelled jury is selected; the first 12 randomly drawn men who have appeared and are not struck for cause form the jury. An individual is struck for cause if there is a clear reason they cannot be impartial, such as knowing the defendant or victim, or if they were not qualified to be in the jury book in the first place. The 1825 Jury Act even describes how to implement randomness. All summoned names

shall be written on a distinct Piece of Parchment or Card, such Pieces of Parchment or Card being all as nearly as may be of equal Size ... and shall ... be put together in a Box to be provided for that Purpose, and when any Issue shall be brought on to be tried, such Associate or Prothonotary shall in open Court draw out Twelve of the said Parchments or Cards one after another, and if any of the Men whose Names shall be so drawn shall not appear, or shall be challenged and set aside, the such further Number, until Twelve Men be drawn, who shall appear, and after all just Causes of Challenge allowed, shall remain as fair and indifferent.

(1825 Jury Act)

If the original jury does not reach a verdict, which must be unanimous, the guidelines indicate that a new jury should be drawn, without returning the original juror cards to the box. Interestingly, the same jury, if not objected to, could try several issues (cases) in succession without being redrawn. In practice, this seems to have occurred quite often.<sup>15</sup> Repeated use of the same jury would potentially limit the degree to which seated jury and case characteristics are related, a problem prevalent in archival analyses of modern day data, particularly in the US. Similarly, if there were any men on the original jury who both parties consented to withdraw, or who was justly challenged by the court, these names could be set aside and the case tried with the residual of the original jury and new jurors were drawn from the box.

Finally, the sheriff maintained records of who had served on juries. Depending on the court and jurisdiction, individuals should not have had to serve again within specified time periods. In practice, however, it is common to see the same men repeatedly serving as jurors.

## 1.2. *The History of Female Jurors*

Historically, only men in England and the US (whose jury system has its origins in English common law) were qualified to be jurors.<sup>16</sup> To justify the exclusion of females from jury service in the 18th and 19th centuries, it was argued that:

<sup>15</sup> This is observed both in our data from the 1920s and even more so in the Proceedings of the Old Bailey, which includes information on the juries trying cases at the Old Bailey from approximately 1750 to 1850.

<sup>16</sup> An exception to this occurred when female defendants sentenced to death pled the belly; in this case, a jury of matrons would examine the women to assess the validity of her claim.

- (i) most women worked domestically and lacked the worldly experience necessary to make informed decisions as jurors and
- (ii) women were too 'delicate' and would be made impure by jury service (Fowler, 2005).

As women's rights movements gained momentum, arguments in favour of female jurors suggested that female jurors would bring their 'domestic virtue' into the courtroom to have a positive influence, increase the representativeness of the community on the jury, and bring a unique perspective into deliberations (Fowler, 2005). Faced with their first case on the exclusion of female jurors, the 1946 Supreme Court decision in *Ballard v. United States* conveyed the idea that each sex contributes something unique to jury deliberations.<sup>17</sup>

In England, the Sex Disqualification (Removal) Act of 1919, which was passed into law on 23 December 1919, enabled women to enter both the legal profession and the civil service and to become jurors.<sup>18</sup> In a broad opening statement, it specified the following:

A person shall not be disqualified by sex or marriage from the exercise of any public function, or from being appointed to or holding any civil or judicial office or post, or from entering or assuming or carrying on any civil profession or vocation, or for admission to any incorporated society ... and a person shall not be exempted by sex or marriage from the liability to serve as a juror.

(Sex Disqualification (Removal) Act of 1919)

In practice, however, females did not begin serving on until January 1921. Correspondence preserved in the National Archives highlights the many practical issues related to implementing this Act, including which females should be in the jury pool, which females could apply for exemption, how the jury would be selected, and whether defendants can complain about the gender of the jury. The statutory rules governing female jurors were passed in October 1920, and had the following general provisions:<sup>19</sup>

- (i) juror books should be updated to include eligible women;
- (ii) jury summons should be independent of gender;
- (iii) the number of women on any panel of jurors should be in (approximately) the same proportion as the number in the jurors' book from which the panel is drawn; and
- (iv) a summoned female can apply for exemption based on pregnancy or other feminine condition;

<sup>17</sup> The decision states 'The truth is that the two sexes are not fungible; a community made up exclusively of one is different than a community composed of both; the subtle interplay of influence one on the other is among the imponderables. To insulate the courtroom from either may not in a given case make an iota of difference. Yet a flavour, a distinct quality is lost if either sex is excluded. The exclusion of one may indeed make the jury less representative of the community than would be true if an economic or racial group were excluded'.

<sup>18</sup> At the same time that females qualified to be jurors in England, similar movements were taking off around the US. (See Fowler (2005) and McCammon (2012) for detailed discussions of the US women's jury movements.) Utah became the first state in 1898 to authorise women to participate on juries, but by the time of the Second World War, 21 states still prohibited female jurors. And three states (Alabama, Mississippi and South Carolina) banned female jurors as late as 1962.

<sup>19</sup> These correspondence are preserved in National Archive document LCO2/559, 'Rules of the Supreme Court (Women Jurors) 1920 made in pursuance of the Sex Disqualification Act 1919'.



- (v) the Under-sheriff has the discretion to exempt any summoned woman with appropriate evidence of pregnancy or other feminine conditions; and finally,
- (vi) the possibility of all female or all male juries was preserved.

The judge could at his discretion (or after a request by the prosecution or defence) require that a jury be all male, all female, or exempt females from service based on the sensitivity of the case.

Newspaper articles shed some light on the attitudes towards this dramatic change. According to a 1920 *New York Times* article, the prosecutor congratulated the women for 'at last taking their proper place in the administration of justice in England'.<sup>20</sup> From the viewpoint of the female jurors, *The Manchester Guardian* (1921) quotes one as saying 'that the presence of women on juries would result in greater care being taken in cases where women were concerned. Some cases would be very unpleasant, but men had not shirked their duties and women must equally show a public spirit'. The article also notes that the London jurywomen 'seemed to have been drawn from all classes, and they varied widely in age'.<sup>21</sup>

Debate regarding both the suitability and potential impact of female jurors continued over the following years. Termed the 'great experiment' in some newspaper headlines, opinions regarding female jurors were greatly varied.<sup>22</sup> A 1926 *Daily Express* opinion column asked: 'Is the woman juror inclined to confuse the issues by the introduction of sympathetic or antipathetic considerations? ... Can woman, in short, suddenly divorce herself from temperamental inconsequence and from the compassionate and emotional instincts on which she often acts, and become in a moment a hard, matter-of-fact analytical administrator of the strict letter of the law?' In response to such comments, Sir Robert Wallace, is quoted as saying that he 'found women excellent as jurors, and has no complaints against them'. In terms of the impact of female jurors on case outcomes, an early assessment reported in an April 1921 *Times* article by the counsel to the Treasury at the Central Criminal Court found that mixed gender juries return the same verdict as a jury of all men. Five years later, opinions may have begun to change. One 1926 newspaper article quotes a court recorder as saying the female jurors are harsher than men: 'I venture to think that a man is safer in the hands of the bench than he would be in the hands of a present-day jury. One finds the fair sex far more severe than the old-fashioned jury'.

### 1.3. Potential Confounding Issues and Mechanisms

The main analysis presented below uses a pre-post research design that compares conviction rates by juries before and after The Sex Disqualification (Removal) Act of 1919 to

<sup>20</sup> According to a *New York Times* article titled 'First women appear on jury in England', the first cases in England to be heard by female jurors occurred on 28 July 1920, when six women formed part of the jury in the Bristol Quarter Sessions. The article goes on to describe how the women heard six cases tried, but at the close of the proceedings, two of the women asked to be excused from further service to care for their children. Two other women replaced them on the jury.

<sup>21</sup> *The Guardian*, Originally published in the *Manchester Guardian* on 12 January 1921, 'The woman juror's new sphere' <http://www.theguardian.com/theguardian/2011/jan/12/archive-the-woman-jurors-new-sphere-1921>.

<sup>22</sup> These newspaper articles, collected by the Home Office, are at the National Archives under HO 45/13321.

study the impact of adding females to the jury pool on conviction rates. This interpretation of the results, however, requires ruling out several potential confounding issues that might invalidate a pre-post design. One potential issue is that the Act not only opened jury service to women (though about half of the original Act is, in fact, dedicated to juries) but also allowed for the inclusion of women in legal service (solicitors and judges) and civil service. While all qualified females were immediately added to the jury books, however, the immediate impact on the civil and legal professions was much more limited. The first four women were admitted as solicitors in 1922 and there were only 100 female solicitors throughout the United Kingdom 10 years later in 1931. The first female court recorders (Rose Heilbron) and judges (Elizabeth Lane) were not appointed until 1956 and 1962, respectively.<sup>23</sup> Thus, the only court participants who were immediately affected by the Act were jurors.

A second potentially confounding issue is that the implementation of the Act itself may have reflected other discrete changes in society regarding gender roles. But the circumstances under which the Act was enacted, and its application in the years to follow, suggest that perhaps these societal changes were not as stark as one might expect. First, the Act was put forth by the governing conservative coalition to counter the minority Labour party's proposed Women's Emancipation Bill, which was much more radical in its proposals to give women equal voting rights and the right to sit and vote in the House of Lords. In presenting the Act to the House of Lords, The Lord Chancellor said that it 'would prove surprising "and to many extremely disagreeable"' (Bennion, 1979). In addition, despite the general language of the Act, it has almost never given rise to litigation; 50 years after the original Act, a Court of Appeals judge said 'This is a most important statutory provision, which, so far as I know, has never yet been considered by the courts' (Bennion, 1979).

The seating of female jurors is the most obvious direct effect of the reform, encompassing a number of potential mechanisms. Female jurors might assess the nature of the evidence or view the defendant, victim, or other court participants differently than men, which might not only affect their own vote(s) but also change the nature of deliberations, and thus the votes of male jurors. A growing body of research evidence points to a range of mechanisms through which gender diversity affects group decisions including both the direct effect of adding new perspectives as well as the more general impact of having more diverse view points on the depth and quality of ensuing discussions (Andreoni and Vesterlund, 2001; Campbell and Mínguez-Vera, 2008; Carter *et al.*, 2010).

It is important to note, however, that our research design identifies the full causal impact of the Act on jury trials, which might not be due completely to the addition of females to the seated jury. Indirect effects of having women in the jury pool might arise for several reasons. First, attention focused on 'striking' female jurors might limit strikes for male jurors, thereby affecting the set of male jurors actually seated. Second, judges might request an all-male jury because of the 'sensitive' nature of the case; a signal about the case that might influence the deliberations and behaviour of the seated male jurors. Finally, media coverage of the 'female juror experiment' could affect the behaviour of men, even when women are not present on the jury. For all of these reasons, it is important to keep in mind that our primary analysis estimates the reduced form

<sup>23</sup> [http://news.bbc.co.uk/2/hi/uk\\_news/40448.stm](http://news.bbc.co.uk/2/hi/uk_news/40448.stm).

effect of the policy rather than the causal effect of seated female jurors on case outcomes *per se*.

In considering potential mechanisms, a final feature of our research design is worth highlighting. In particular, the existing literature on the effect of jury gender composition has been based primarily on modern criminal trials in the US, where the variation in female representation generally falls in a somewhat narrow range around 50%. Thus, setting endogeneity concerns aside, another potential reason that the previous literature might find little effect of jury gender composition is that increasing female juror representation from 40% to 60% may not be a particularly impactful margin. In contrast, one might expect that increasing female representation from zero to just one or two jurors might have a more significant effect, as the presence of even a single female juror might substantially change the nature of the deliberations and decisions, especially when verdicts must be unanimous (as in our study). A unique feature of our study relative to the previous literature is that we identify the impact of female jurors precisely at this margin of inclusion, a margin that remains policy relevant around the world and in many institutions, such as the US military.<sup>24</sup>

## 2. Data

### 2.1. Data Description

We collected data for this project from the Central Criminal Court: Court Books, which are housed in England's National Archives. These books are original hand-written documents by court reporters that have summary details for each trial, including the date(s), judge name, number of charges, the main offences, plea, verdict, sentence, juror names, defendant names and typically victim names. We obtained photocopies of the complete Court Books for the First Court (from October 1917 to October 1926) and the Second Court (from January 1919 to January 1925); this amounted to about 3,500 pages of case summaries from 193 court sessions. See Appendix Figure A1 for a photograph of an example of a jury trial in Court 1. We manually transcribed these records into a defendant by case level data set, including all cases for which there was a seated jury. For the first court, we also coded cases in which the defendant confessed and there was no seated jury. For cases with multiple defendants, we keep track of both the case number and the defendants' order; generally, the charges associated with the first defendant are more serious than those associated with the last. This resulted in a data set of 3,335 defendant-case observations (1,080 jury trials and 1,187 confessions in the First Court and 1,068 jury trials in the Second Court).

From this raw data, we create a number of variables to characterise the defendant and case, as well as the jury. With respect to gender, the focus of this paper, we identify juror gender, victim gender and defendant gender on the basis of names. We focus on first names, but in instances when the gender is ambiguous, we rely on middle names, which are almost always provided. For victim gender, when there is no name provided, we code

it as unknown or none (meaning no person), but it could be the government or a firm that was the victim.

<sup>24</sup> To our knowledge, there is no data available that would allow one to study a modern setting (such as the US military) where the key variation in women in the jury pool is none to a few.

We then use the gender-classified juror names to identify the number of females seated on each 12-person jury; unfortunately, the only thing we know about seated jurors is their names. However, we do create a jury identifier to keep track of whether the same jury tried multiple cases: the court documents often say 'same jury sworn'. In fact, we observe 1,047 unique juries in the data, but more than 2,000 defendant-trial observations.

We identify the total number of charges and classify each offence (of which there were more than 100 detailed offence types listed) into one of 15 broad categories: killing (manslaughter, murder and infanticide), aggravated assault, arson, robbery, burglary, rape, theft, bigamy (i.e. the act of marrying while already married), deception, offences against the king, other offences against persons, other property offences, other sex offences, conspiracy and all other offences.<sup>25</sup> Attempts (and assisting) to commit these offences are included in the same categories. We total the number of (known) charges in each broad category, and create dummies indicating whether the defendant is charged with at least one offence in each category. While the court records almost always indicate the total number of charges, and the types of offences for the first charge, the specific offences of the later (typically less serious) charges are not always listed. We also identify whether the offence was against a child, and whether the defendant was charged with having a criminal history.

We also code the verdict for each offence: whether the defendant confessed (in which case there is no jury verdict) or whether the case was *put* to the jury. In most jury trials, the jury returned a verdict of guilty or not guilty. However, we also note those cases in which a jury was seated but never delivered a verdict, either because it was a hung jury, a directed verdict or a confession/plea in the middle of trial. For cases with a jury verdict, we identify whether the first offence (generally the most serious offence) resulted in conviction as well as the share of total charges facing the jury that resulted in conviction.

Additional variables that we code are the length of the trial in days, which is suggestive of the quality of evidence and the seriousness of the case, and whether the defendant was undefended. Attorney names were typically listed in the margins, but in most instances, they were too hard to read to confidently code; 'undefended' was clearly indicated however. Though juries only decided verdicts, we also code the sentences associated with each case. The most common sentences include prison (with or without hard labour), penal servitude (which was also a form of imprisonment with hard labour) and death.

A close reading of these historical court records also provides some data that can be used in a more qualitative manner to understand the jury system at the time and the role of female jurors. At the beginning of many court sessions, for instance, a list

<sup>25</sup> Some offence categories include multiple sub-categories. Killing includes murder, manslaughter, child murder and infanticide. Rape includes charges of rape as well as indecent assault. Burglary includes burglary as well as having the tools for burglary. Theft includes auto theft, theft, postal theft, larceny, blackmail, embezzlement, bribery, shoplifting and extortion. The main 'other property offense' is receiving stolen goods. 'Other offenses against persons' include assault, threatening behaviour, kidnapping, and libel. 'Other sex offenses' include buggery, pornography, possession of indecent print, sending indecent cards, indecent exposure, gross indecency, incest, procurement and prostitution. Deception primarily includes forgery, fraud and perjury. Offences against the king include coining, tax fraud, seditious speech, treason, trading with the enemy, blasphemous libel, communicating secret information, conspiracy against the military and uttering. Finally, the all other offence category includes offences such as abortion, child neglect, suicide, concealing birth and firearms.

of jurors who are fined (for not showing up at court) was provided, including the juror name and profession. Females do not show up in these lists very often (just 17 instances): 75% of those that do are characterised as married, widowed, a spinster, a gentlewoman and a householder. The remaining are an artist, a boarding house keeper, a mathematical coach and a shipping agent. In contrast, male professions range from professional (accountants, stockbrokers, engineers), to retailers, manufacturers, manual labourers and artists; more than 10%, however, are classified as gentlemen. In addition, for a handful of cases (11), there was a note in the documents that the judge requested an all-male jury; it is possible, however, that such requests were not always noted in the records. Finally, we also see notes in the margins by the court reporter indicating if the jury was provided certificates for lunch or tea, and even hotel lodging, as well as whether the jury was exempted from future jury service beyond the normal amount. While we have no way of knowing whether this information was always noted by the court reporter, we code it as a way of possibly controlling for the most serious of cases.

## 2.2. *Summary Statistics*

**Table 1** presents summary statistics for jury trials overall, and broken down by court. The statistics for offence categories reveal that 20% of defendants in jury trials have at least one charge of killing, the most serious of offence categories. 5% of defendants are charged with robbery while 8% are charged with aggravated assault and 8% with another offence against a person (i.e. violent in nature). Among property crimes, 10% are charged with burglary, 14% with theft and 15% some other property crime. For sex offences, 6% are charged with rape and 9% with another sex offence. Finally, among those offences that do not easily fall into the categories of violent, property and sex offences, 3% of defendants are charged with bigamy, 16% with deception, 3% with an offence against the king and 6% with another offence. It is also worth noting that 14% are identified as having a criminal history and the average number of charges in total is 3.2.

The next panel of **Table 1** reports statistics for defendant, victim and other case characteristics; 13% of defendants are female and there are, on average, 1.9 defendants per case. 25% of the cases are identified as having a female victim and 40% as having a male victim and 36% of the cases are classified as having no victim.<sup>26</sup> 4% of the cases have a child victim and 21% of defendants are undefended. The average trial length is 1.9 days, although this does not mean that the trial actually took two complete days but rather that the court was in session, on average, on two dates.

Turning to case outcomes, of the 2,148 trials for which a jury was seated, 3% resulted in a hung jury on at least one offence and 8% were dismissed by the judge for any reason (including a hung jury) before the jury made a verdict (it is unclear at what stage in the trial some of these cases were dismissed). For the cases in which the jury returned a verdict on at least one charge, 60% of defendants are found guilty on at least one charge, 45% of defendants are found guilty of all charges and 49% are found guilty on

<sup>26</sup> This classification is difficult from a modern viewpoint, when considering offences like abortion or burglary. Also note that there are some missing observations for victim characteristics due to no names being listed in the court summaries or difficulties reading the names (or assigning gender).

Table 1  
Summary Statistics for Cases with a Seated Jury: Overall and by Court

Variable	Both Courts			Court 1			Court 2		
	Observations	Mean	SD	Observations	Mean	SD	Observations	Mean	SD
post_reform (post-1921)	2,148	0.63	0.48	1,080	0.67	0.47	1,068	0.59	0.49
any_female_jurors	2,147	0.49	0.50	1,079	0.52	0.50	1,068	0.46	0.50
No. female jurors	2,147	1.29	1.58	1,079	1.28	1.46	1,068	1.30	1.69
<i>Offence characteristics</i>									
No. counts	2,148	3.22	5.39	1,080	3.20	4.64	1,068	3.25	6.07
<i>violent</i>									
any_killing	2,148	0.20	0.40	1,080	0.39	0.49	1,068	0.02	0.13
any_aggravassault	2,148	0.08	0.27	1,080	0.10	0.30	1,068	0.06	0.24
any_arson	2,148	0.02	0.13	1,080	0.03	0.16	1,068	0.01	0.09
any_robbery	2,148	0.05	0.21	1,080	0.02	0.15	1,068	0.07	0.26
any_other_off. <i>versus</i> person	2,148	0.08	0.27	1,080	0.09	0.29	1,068	0.06	0.24
<i>property</i>									
any_theft	2,148	0.14	0.35	1,080	0.05	0.21	1,068	0.23	0.42
any_burglary	2,148	0.10	0.30	1,080	0.06	0.23	1,068	0.15	0.35
any_other_property	2,148	0.15	0.35	1,080	0.04	0.20	1,068	0.25	0.43
<i>Sex offence</i>									
any_other_sex offence	2,148	0.09	0.28	1,080	0.05	0.22	1,068	0.13	0.33
any_rape	2,148	0.06	0.24	1,080	0.08	0.27	1,068	0.05	0.22
<i>Other offence</i>									
any_deception	2,148	0.16	0.36	1,080	0.13	0.34	1,068	0.18	0.38
any_bigamy	2,148	0.03	0.17	1,080	0.01	0.12	1,068	0.05	0.21
any_offence_against_king	2,148	0.03	0.18	1,080	0.04	0.20	1,068	0.03	0.16
any_other	2,148	0.06	0.24	1,080	0.10	0.31	1,068	0.01	0.11
charged_with_crim_history	2,148	0.14	0.35	1,080	0.08	0.28	1,068	0.21	0.40
<i>Defendant and victim characteristics</i>									
def_female	2,131	0.13	0.33	1,070	0.17	0.37	1,061	0.08	0.28
No. defendants	2,148	1.86	1.48	1,080	1.89	1.70	1,068	1.84	1.23
female_victim	2,051	0.25	0.43	1,011	0.34	0.48	1,040	0.15	0.36
male_victim	2,051	0.40	0.49	1,011	0.39	0.49	1,040	0.42	0.49



Table 1  
(Continued)

Variable	Both Courts			Court 1			Court 2		
	Observations	Mean	SD	Observations	Mean	SD	Observations	Mean	SD
no_victim	2,051	0.36	0.48	1,011	0.28	0.45	1,040	0.43	0.50
child_victim	2,148	0.04	0.20	1,080	0.07	0.25	1,068	0.02	0.14
trial_length (days)	2,142	1.91	2.10	1,079	1.87	1.80	1,063	1.94	2.37
Undeclared	2,148	0.21	0.41	1,080	0.16	0.36	1,068	0.27	0.44
<i>Case Outcomes</i>									
Guilty_any offence (1/0)	2,027	0.60	0.49	998	0.68	0.47	1,029	0.53	0.50
guilty_share offences	2,021	0.53	0.46	992	0.61	0.46	1,029	0.46	0.46
guilty_all offences (1/0)	2,021	0.45	0.50	992	0.54	0.50	1,029	0.37	0.48
guilty_1st offence (1/0)	1,972	0.49	0.50	968	0.57	0.50	1,004	0.42	0.49
jury_hung	2,148	0.03	0.16	1,080	0.02	0.15	1,068	0.03	0.17
jury_dismissed or hung	2,148	0.08	0.28	1,080	0.08	0.27	1,068	0.09	0.28
Prison	1,286	0.63	0.48	744	0.53	0.50	542	0.77	0.42
Penal servitude	1,286	0.20	0.40	744	0.25	0.43	542	0.14	0.35
Death	1,286	0.04	0.20	744	0.07	0.25	542	0.00	0.04
other_sentence	1,286	0.12	0.33	744	0.15	0.36	542	0.08	0.27

Notes. Sample sizes vary due to the following:

- (i) difficulty identifying names (or associated gender),
- (ii) incomplete information in the original case summaries, and
- (iii) different subsamples.

Sentencing variables are only defined for cases that resulted in at least one guilty verdict. Jury hung and dismissed variables are based on cases that had a jury seated, regardless of whether that jury reached a verdict. The guilty variables are based on the sample of cases for which at least one charge had a verdict decided by the jury; but, even this can vary across outcome variables since for instance, the first charge may not have been put to the jury.

the first (typically most serious) charge. The average share of charges on which a guilty verdict is returned is 53%.<sup>27</sup> For those found guilty, we also report judge sentences: 63% are sentenced to prison, 20% to penal servitude, 4% to death and 12% to some other punishment such as a fine or corporal punishment.

Table 1 also clearly demonstrates the differences between the First and Second Court. Broadly speaking, cases in the First Court are more serious and can be eligible for capital punishment. 39% of the cases in the First Court have at least one killing charge, compared to 2% in the Second Court. Likewise, rape and aggravated assault are more prevalent in the First Court, while the Second Court largely sees various types of property crimes. An examination of the sentencing data also indicates that the crimes in the First Court are more serious, and eligible for harsher punishment. 7% of the defendants receive a death sentence in the First Court compared to none in the Second Court. A higher share of defendants in the Second Court are sentenced to prison than the First Court (77% *versus* 53%) while a higher share in the First are sentenced to penal servitude, which is harsher than prison.

Conviction rates also differ across the two courts. For instance, 54% of defendants in the First Court are found guilty of all charges compared to 37% in the Second; likewise, 68% in the First are guilty of at least one charge compared to 53% in the Second. These differences in conviction rates likely reflect differing offence distributions, and the quality and nature of evidence for various offences. Other differences across courts include a higher share of undefended cases in the less serious Second Court, and differing defendant/victim characteristics. The latter is unsurprising, given the different offences.

### 3. The 'First Stage' Impact of the Reform

#### 3.1. *The Effect of the Reform on the Presence of Females on the Jury*

This Section characterises the implementation of the reform. We begin by examining when females were actually seated on juries in Figure 1, which presents annual data for all cases in which a jury was seated. The top panel considers all cases and presents the extensive margin (whether there were any females seated) on the left and the intensive margin (the number of females seated) on the right. Consistent with the historical documentation described in Section 1, no females were seated on London juries prior to 1921. But, in 1921, more than 60% of juries had at least one female and the average number of females seated was about 1.7 (note that this includes cases with zero females). The presence of females on the jury continued to increase over subsequent years so that by 1923 about 90% of cases had a female juror and the average number of female jurors was almost 2.5.

Table 2 quantifies the first stage impact of the reform, presenting regressions of whether any females were seated and the number of females seated on dummies for

<sup>27</sup> There is some variation in sample sizes across outcome variables depending on which charges were put to (and decided by) the jury and whether any information was omitted from the records. In addition, cases can fall into multiple classifications. At the most extreme, consider a hypothetical defendant charged with five offences, to which he confesses to two. Three charges are put to the jury – the jury reaches a verdict for two. The jury may be dismissed from deciding the final charge (hung, are directed by the judge, or the defendant confesses).

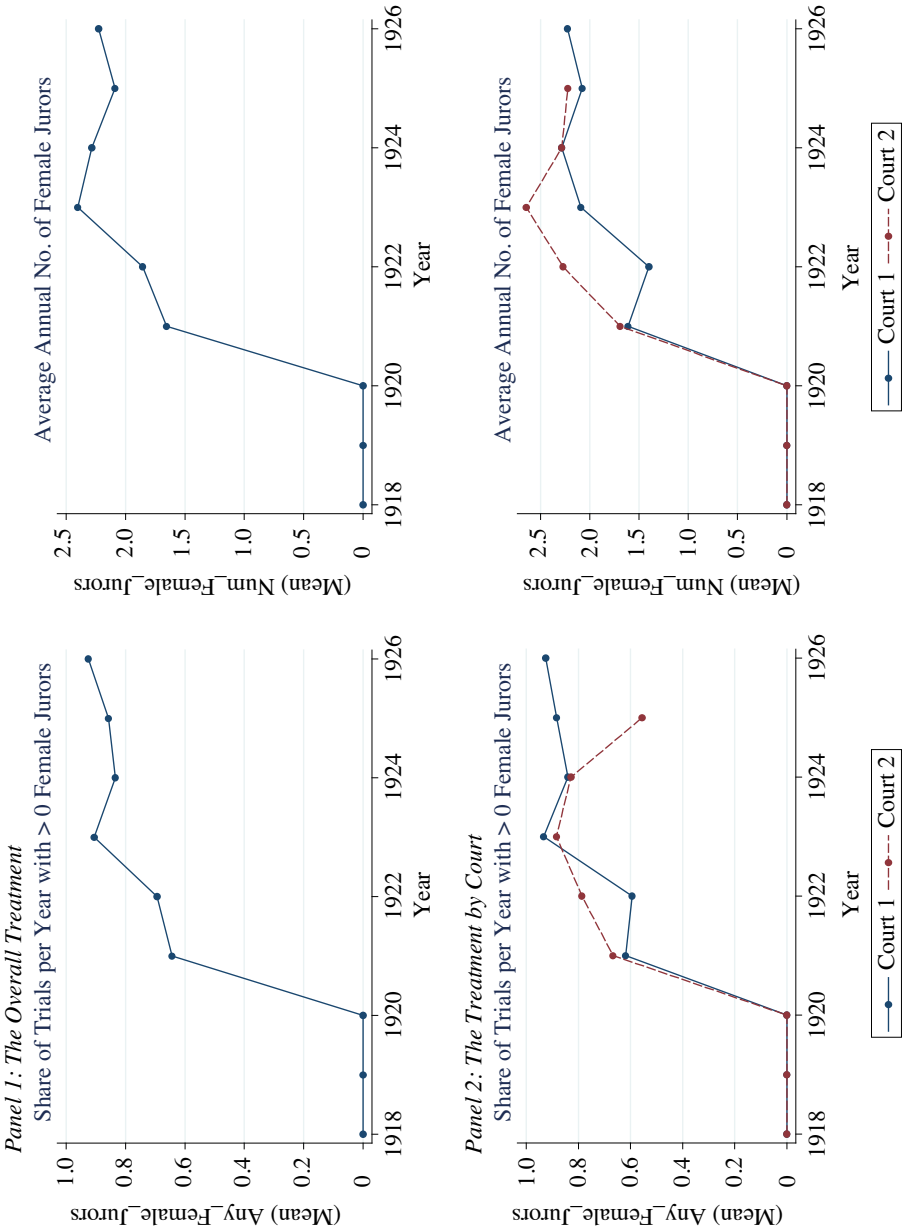
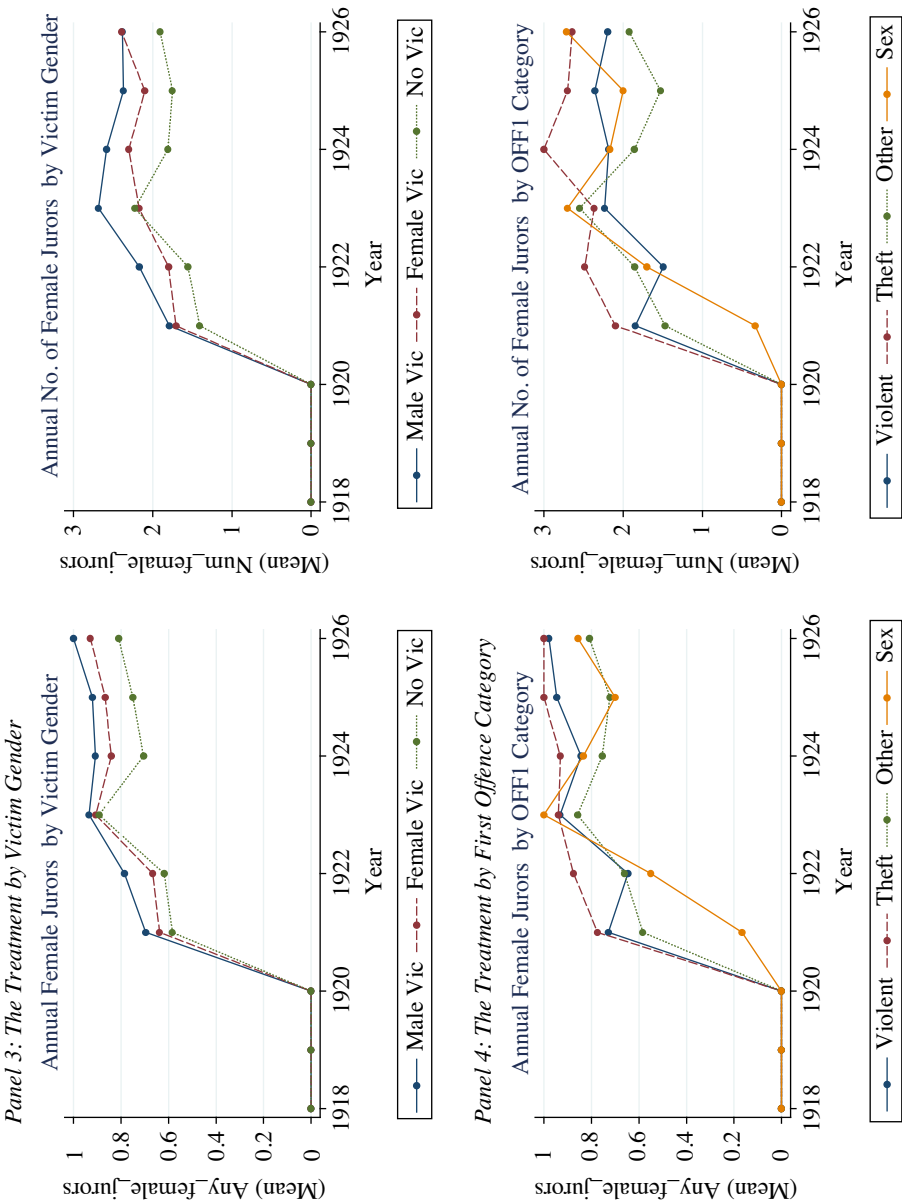


Fig. 1. *The Treatment: Share of Cases with Female Jurors and Average Number of Seated Jurors*  
Note. Colour figure can be viewed at <https://academic.oup.com/ectj>.



each post-reform year, from 1921 to 1925; the 1925 dummy also includes a small sample of cases tried in 1926. The estimated impact of the reform on female representation on juries is independent of controlling for a large set of offence and case characteristics, as can be seen by comparing the even and odd numbered columns in Table 2. In the first year of the reform, the likelihood of having a female juror increases by 64 percentage points and there is an average of 1.6 female jurors seated. It appears to have taken about three years to reach full implementation and a steady state level of female representation on jurors.

The second panel of Figure 1 assesses whether these ‘first stage’ effects of the reform on female representation were heterogeneous across the First and Second Courts. At both the extensive and intensive margins, the effects of the reform are comparable in the first year (1921) across courts. Approximately, 60% of cases have at least one female seated and on average, there are about 1.6 females seated (including all male juries). While a secondary increase in the use of female jurors in the next two years is observable in both courts, it appears to occur faster in the Second Court. The presence of female jurors on the First Court holds about constant from 1921 to 1922 at both margins. In the Second Court, the share of cases with female jurors increases already to almost 80% in 1922 and the average number of seated jurors increases to 2.3 in 1922 and 2.6 in 1923.

The previous figures have only displayed the average number of seated females on the jury. An examination of the distribution of seated females (not shown) shows that more

Table 2  
*‘First Stage’ Effect of Reform on Likelihood of Female Juror*

	(1)	(2)	(3)	(4)
	Dependent variable:			
	any_female_jurors		num_female_jurors	
year_1921	0.644*** (0.021)	0.627*** (0.022)	1.654*** (0.082)	1.618*** (0.084)
year_1922	0.694*** (0.021)	0.696*** (0.021)	1.855*** (0.082)	1.847*** (0.082)
year_1923	0.905*** (0.022)	0.894*** (0.022)	2.404*** (0.085)	2.362*** (0.086)
year_1924	0.834*** (0.023)	0.803*** (0.023)	2.286*** (0.087)	2.169*** (0.087)
year_1925	0.888*** (0.025)	0.885*** (0.026)	2.151*** (0.095)	2.245*** (0.100)
_cons	8.72e-15 (0.011)	0.0649* (0.036)	1.49e-14 (0.043)	0.263* (0.137)
Court	No	Yes	No	Yes
Offence categories	No	Yes	No	Yes
Other case characteristics 1	No	Yes	No	Yes
Month dummies	No	Yes	No	Yes
Other case characteristics 2	No	Yes	No	Yes

Notes. Estimated with OLS. Other case characteristics 1 include victim characteristics (male, female, no and child), defendant gender, number of defendants, codefendnat id, no. charges and charged with criminal history. Other case characteristics 2 include trial length, undefended and jury breaks.

than 30% of cases in the first two post-reform years have no seated females. Almost 15% have one female, more than 20% have two females, and about 15% have three females. More than 10% of juries contain four or five seated females. Just a handful had more than five females, and there were never more than seven in any year. In the following years, there is a sharp increase in the share of juries with two females (30%), three females (almost 25%) and four females (about 12%).

Our analysis of seated jurors provides several key facts about female jury service following the reform:

- (i) despite passing in 1919, the reform was not implemented until 1921;
- (ii) in the first post-reform year, 1921, a female was seated on more than half of the juries;
- (iii) the reform gained further traction in the two subsequent years, substantially increasing the presence of female jurors; and
- (iv) while males clearly still dominated the jury, more than 20% of post-reform juries included at least three females among the 12 jurors.

### 3.2. *Post-reform Selection of Female Jurors*

Having shown that women were indeed seated on juries following the reform, we now turn to the question of whether they were systematically more likely to be challenged or removed from certain cases on the basis of observable characteristics, such as offence type or victim gender. Such systematic deselection would be consistent with anecdotes from newspaper coverage at the time. A 1921 *Times* article describes a challenge by the defence against two female jurors on an indictment for an offence against a girl aged 11 and three jurywomen on another case of a similar nature. Another *Times* article highlights the ‘practice of challenging women on juries in cases of a sexual character, in which young girls were concerned’ while another quotes a judge who gave the female jurors the option of leaving (at any attorney request) on a case involving relations between two men.<sup>28</sup>

To begin to address this question, the third panel of [Figure 1](#) breaks out cases according to whether they had at least one female victim, male victim or no easily identifiable victims. The same general time pattern in the use of female jurors is observed across each

<sup>28</sup> An article titled ‘Women jurors in unpleasant cases’ quoted the judge: ‘The question at issue in this case involves relations between two men. It will involve sexual points of the gravest indelicacy – questions which even men would hesitate to discuss amongst themselves. If the ladies of the jury, or any one of them, desires to remain in that body so be it, but if the ladies prefer that their places be taken by men I will assent to their request’. One of three female jurors left the court.



subcategory. While female jurors are slightly more likely to be present in cases involving male victims, these gaps are quite small, however, and not generally statistically significant.<sup>29</sup> Decomposing all post-reform cases by victim gender reveals that female jurors are seated more often for cases involving male *versus* female victims (84% *versus* 79%) – see Appendix Table B1.

The last panel of Figure 1 classifies the cases according to whether the first charge is a violent offence, property crime, sex offence or other offence. Though the same general time pattern is again seen across all cases, female presence on the jury was strongest for property crimes and weakest for sex offences, especially in the years immediately following the reform. In particular, while less than 20% of sex offence cases had a female juror in the first year following the reform, by 1923, the presence of females on sex offence juries was comparable to that for other offence types, as can be seen in Figure 1.

When looking across crime categories over the entire post-reform period in Appendix Table B1, we see that at least 75% of trials have female jury representation, with an average number of females seated of 1.8 or more. The only exceptions are sex offences: for cases with at least one rape charge, just 56% of trials have a female juror, and the average number seated is 1.6. The presence of females is even lower for the other sex offence category, which includes offences such as buggery, incest, pornography, indecent exposure and gross indecency: just 27% of cases have a female juror and the average number of female jurors is 0.7. Thus, there is clear evidence that females are deselected from sex offence cases, especially in the two years immediately following the reform.

Appendix Table B2 presents regressions of whether any females are seated on a court dummy, offence characteristics, victim and defendant characteristics, and other case characteristics. Even in this multivariate regression framework, there is significant selection of females based on offence type. Not surprisingly, females are less likely to be seated on rape and other sex offence cases; this systematic selection for rape cases is only seen during the first three years. Having controlled for case characteristics (namely offence type), however, female jurors are not seated differentially on female *versus* male victim cases. However, there is some evidence that females are significantly less likely to be seated on ‘no victim’ cases. Again, it is important to note that ‘no victim’ cases are quite heterogeneous in nature, and can include offences against firms or the government rather than people, or offences such as buggery which in many instances would be seen as victimless in modern definitions. There is no significant effect of defendant gender on the female presence on the jury.

Taken as a whole, these results suggest that there is considerable variation in the seating of female jurors on the basis of observable (and presumably unobservable) case characteristics. Such concerns about the non-random seating of female juries motivate our primary pre-post research design as well as the other complementary analyses that we describe in the next Section.

<sup>29</sup> Likewise, Appendix Table B1 demonstrates females are more often seated on cases with a female *versus* male defendant (85% *versus* 77%).

## 4. Estimating the Effect of the Reform on Case Outcomes

### 4.1. Methodology/Identification Strategy

This Section presents the results from a pre-post design that compares conviction rates by juries before and after The Sex (Disqualification) Act of 1919 in England. Specifically, our research design will compare conviction rates prior to the implementation of the Act in 1921 (when there were no female jurors in the pool) to conviction rates from 1921 on, when female jurors were regularly observed in the pool and seated juries. This research setting allows us to study the role of female jurors using real world data in a way that overcomes the fundamental challenge that has arisen in prior research that examined the impact of the gender composition of the seated jury. Seated juries in the US and UK (though to a lesser extent) result from a de-selection process that strikes potential jurors on the basis of observable (and unobservable) characteristics. Even though one can no longer legally strike jurors on the basis of gender, there may be other justifications that are in fact correlated with gender. Thus, seated juries are not random and the characteristics of the jury can conceivably be related to unobserved case factors such as the amount of evidence present. As a result, a causal interpretation cannot reliably be attached to the results of studies that only consider the variation in seated jury composition on trial outcomes. In contrast, our research design exploits exogenous variation in the gender composition of the seated jury due to a law change.<sup>30</sup>

Because we identify the causal impact of the reform using a simple pre-post design (without control groups), it is important to consider the time windows that make this design most accurate. The further observations get from the date of implementation, the less clean the design becomes, as other things change in society, the economy and the justice system. On the other hand, the further away we get from implementation, the more observations that are available, potentially increasing the precision of the analysis. In an effort to balance these two issues, we focus on a three-year window around the reform in our baseline specifications, thereby comparing outcomes in the first three years after the reform (1921–3) to outcomes three years before the reform (1918–20). However, to increase the precision of the other parameter estimates, sample observations for all years (1918–26) are included in the analysis, as well as a dummy variable to capture this additional post-reform period. We chose a three-year window as our baseline because the reform is fully implemented by this time, and it is symmetric with the amount of pre-reform data available. In additional analyses presented below, we also demonstrate that the main results are not sensitive to the choice of window.<sup>31</sup>

The equation below presents the baseline specification, where the dependent variable is some measure of whether defendant  $i$  on case  $j$  tried on date  $t$  is convicted by the

<sup>30</sup> Another way prior literature has dealt with this endogeneity issue is to take advantage of the exogenous day to day variation in the jury pool (Anwar *et al.*, 2012, 2014). This approach is not used here as we do not observe the composition of the jury pool.

<sup>31</sup> The main findings are also robust to starting the data in 1919, after the end of the First World War. During the war, anecdotal evidence suggests that the jury pool was different and that selection into the pool did not strictly adhere to regulations.

jury:

$$G_{ijt} = \alpha + \beta_0 \text{PostRef3}_{ijt} + \beta_1 \text{PostRef4plus}_{ijt} + \text{Ct2}_{jt} + \text{Off}_{ijt} \pi + \mathbf{X}_{ijt} \delta + \varepsilon_{ijt}. \quad (1)$$

The primary-dependent variable ( $G$ ) is whether the defendant is convicted of at least one offence put to the jury; robustness checks consider the share of convictions, convictions on the first offence and whether convicted on all charges. *PostRef3* is a dummy for the first three years after the reform, while *PostRef4plus* captures all additional post-reform years.  $\beta_0$  provides the cleanest estimates of the reduced form effect of the reform. Given that the types of cases are different across courts, all specifications include a court dummy (*Ct2*). *Off* is a vector of 15 dummy variables characterising the broad offence categories of the charges.  $\mathbf{X}$  is a vector of case characteristics, including victim gender, defendant gender, number of charges, number of defendants, defendant ranking order and child victim.

#### 4.2. Identification Concerns

Given that we are estimating the pre-post impact of the reform on conviction rates, it is important to ask whether anything else changed at the time of the reform, which would contaminate the pre-post design. Table 3 takes a simple first look at this issue by comparing the sample characteristics for jury trials before and after the reform. In the First Court, we see little change in terms of the offence distribution and defendant/case characteristics. In the Second Court, there are a number of statistically significant differences in case characteristics, although these changes are generally small in magnitude. In the regressions that follow, we control directly for all observable offence and case characteristics but it is important to point out that we cannot rule out the possibility that some omitted case characteristics changed at the time of the reform.

While the differences in the types of offences seen after the reform (in the Second Court) could be due to a change in the types of offences committed (or emphasised by the justice system) at this time, it could also be that a change in plea behaviour results in a different composition of cases reaching a jury trial. If attorneys have perceptions about female jurors' likelihood to convict, then defendants may be more or less likely to plea after the reform. Similarly, it is also possible that prosecutors' perceptions of the likelihood of obtaining a guilty verdict for certain types of offences or that victims' reporting behaviour of crimes changes with the presence of females on the jury.<sup>32</sup> All of these channels (i.e. prosecutor decisions, defendant plea decisions and victim reporting decisions) could theoretically result in a change in the composition of the cases actually brought to trial, thereby affecting the average likelihood of conviction. We take two steps to address these concerns. First, we directly examine the First Court jury trial and plea data to test directly for whether there is a change in plea behaviour.<sup>33</sup> Second, we use sentence length as a proxy for all observable and unobservable case characteristics, and

<sup>32</sup> For instance, Iyer *et al.* (2012) provide evidence in India that mandated representation of females in government increased the reporting of crimes against females.

<sup>33</sup> Generally, plea cases are more minor; defendants tend to be charged with property rather than violent offences, they do not/cannot result in a death sentence, and there are fewer total charges. A larger share of pleas are made by defendants that are undefended (57%) compared to trial defendants (16%).

Table 3  
*Comparison of Pre and Post-Case Characteristics by Court (for Jury Trials)*

	Court 1 jury trials		Court 2 jury trials	
	Pre-1921 N = 355	Post-1921 N = 725	Pre-1921 N = 443	Post-1921 N = 625
undefended	0.15	0.16	0.25	0.28
def_female	0.15	0.17	0.08	0.08
No. defendants	1.80	1.93	1.94	1.77*
female_victim	0.38	0.33	0.12	0.18*
male_victim	0.38	0.39	0.41	0.43
no_victim	0.26	0.29	0.47	0.40*
child_victim	0.06	0.07	0.02	0.02
No. counts	3.00	3.30	2.48	3.79*
any_killing	0.42	0.38	0.00	0.02*
any_aggassault	0.08	0.10	0.05	0.07
any_arson	0.02	0.03	0.00	0.01*
any_robbery	0.03	0.02	0.06	0.08
any_other_off. versus person	0.09	0.10	0.04	0.08*
any_burglary	0.06	0.06	0.17	0.13*
any_theft	0.04	0.05	0.26	0.22
any_other_property	0.03	0.04	0.29	0.23*
any_other_sexoffence	0.06	0.05	0.16	0.10*
any_rape	0.08	0.07	0.04	0.05
any_deception	0.14	0.13	0.13	0.21*
any_offence_against_king	0.06	0.03*	0.02	0.03
any_bigamy	0.01	0.01	0.05	0.05
any_other	0.09	0.11	0.00	0.02*
charged_with crim history	0.06	0.09	0.17	0.23*

Note. \*Indicates pre-post difference is significant at 5%.

assess whether the average sentence length changed after the reform for the sample of plea cases.

Table 4 considers whether there is a change in plea behaviour by regressing whether the defendant pled guilty to all charges on our three-year treatment window dummies. When not controlling for anything else, column (1) demonstrates that defendants are 12 percentage points less likely to plea after the reform. But when just controlling for a vector of offence types in column (2), the post-reform plea effect goes to zero, suggesting that conditional on the type of offence charged, there is no change in the likelihood to plea as a result of the reform. The significant effect on the likelihood of plea behaviour in column (1) is driven by a change in offence composition seen in the courts (in particular, the share of bigamy cases), and not a change in the likelihood of pleading within a given offence category.<sup>34</sup> In columns (3)–(6), we also see that there is no significant effect on the likelihood of pleading when decomposing the data into broad offence categories: violent, property, sex and other. Results not shown also indicate that there is no change in the likelihood of pleading after the reform for female victim or female defendant cases. The fact that a defendant's decision to go to trial did not change in response to

<sup>34</sup> More specifically, it is driven by a substantial reduction in the share of bigamy cases seen in the court; there was a gradual decline, which began before the reform, from more than 60% of plea cases in 1918–9 to less than 20% by 1922. This is likely to be explained by the distance from the war. Just controlling for bigamy in column (1) is enough to eliminate the significant coefficient.

Table 4  
*Does Plea Behaviour Change after the Reform?*

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable = plea to all charges					
	All cases	Violent	Property	Sex	Other	
Post-reform three years	−0.115*** (0.025)	0.002 (0.020)	0.010 (0.031)	−0.009 (0.054)	0.059 (0.077)	−0.036 (0.030)
_cons	0.526*** (0.017)	0.557*** (0.029)	0.345*** (0.050)	0.986*** (0.093)	0.561*** (0.137)	0.415*** (0.080)
Offence type dummies	No	Yes	Yes	Yes	Yes	Yes
Other case characteristics	No	No	No	No	No	No
N	2,267	2,267	704	468	201	952
R <sup>2</sup>	0.009	0.397	0.089	0.223	0.109	0.402

Notes. Robust standard errors in parentheses. \*10%, \*\*5%, \*\*\*1%. Note that these specifications are restricted to the First Court, where the plea and trial data have been recorded. Post-reform three years is a dummy equal to one for the first three years after the reform: 1921, 1922 and 1923. All specifications also control for a dummy after the treatment period: 1924–6.

Table 5  
*Sentence Length Analysis: Did Defendant Plea and Prosecutor Charge Decisions Change Case Composition?*

	(1)	(2)	(3)	(4)
	Sample = plea guilty to all charges (court 1)			
	Dependent variable = months prison or penal servitude:			
	Excluding those with other sentences	Assigning 0 to other sentences (fines, probation)		
Post-reform three years	1.726 (1.272)	−1.313 (1.210)	0.946 (1.14)	−1.539 (1.091)
Constant	13.33*** (0.81)	14.87*** (2.33)	11.35*** (0.74)	10.76*** (2.18)
Offence type controls	No	Yes	No	Yes
Other controls	No	No	No	No
N	851	851	1,003	1,003
R <sup>2</sup>	0.005	0.185	0.005	0.174

Notes. Robust standard errors in parentheses; \*10%, \*\*5%, \*\*\*1%. Post-reform three years is a dummy equal to one for the first three years after the reform: 1921, 1922 and 1923. All specifications also control for a dummy after the treatment period: 1924–6.

females being added to the jury pool suggests that one can look at conviction rates to identify the true impact of the reform, rather than accounting for decisions at earlier stages of the judicial process.<sup>35</sup>

Table 5 provides additional evidence that the reform did not change the composition of cases to which a defendant pled guilty. Specifically, columns (1)–(4) of

<sup>35</sup> Robustness checks presented in Appendix Table B5 and B6 confirm that our main results are not sensitive to including plea cases for the first court in the regressions and assigning them a guilty verdict.

**Table 5** consider the sample of cases in the First Court to which the defendant pled guilty on all charges and regresses sentence length (measured in months sentenced to prison or penal servitude) on our three-year treatment window dummies. Columns (1) and (2) exclude those individuals (about 15% of the sample) who were not sentenced to prison or penal servitude (and were given a non-custodial sentence, such as a fine) while columns (3) and (4) assign a sentence length of zero months to those who received a non-custodial sentence. Regardless of how we treat non-custodial sentences or whether any controls for offence types are included in the specification, there is no evidence that the sentence length given by judges to defendants who pled guilty significantly changed after the reform. Taking sentence length as a proxy for both observable and unobservable case characteristics, the lack of a systematic change in sentence length for plea cases is suggestive that there was no systematic change in the composition of observable and unobservable case characteristics as a result of the defendant's plea decision, the prosecutors' charge decisions or the victims' reporting decisions.<sup>36</sup>

#### 4.3. *Pre-post Results*

We now present the results of the pre-post reform analysis, beginning with estimates of the overall effect of the reform on conviction rates. The above discussions indicate the importance of conditioning the analysis on offence type. Thus, our baseline specification includes the vector of offence categories, as well as a court indicator (which to some extent also captures the type of offence). **Table 6** begins in column (1) by looking at the effect of the reform on the likelihood of being convicted of at least one offence with these baseline controls included, revealing that there is not a significant impact of the reform on the likelihood of conviction when aggregating all cases together.<sup>37</sup> There are many features of criminal cases, however, that may be more or less 'salient to females', including the nature of the offence and the gender of the victim and defendant. Thus, in the remainder of our analysis, we examine the heterogeneity in the impact of the reform in cases that may be more or less female salient, respectively.

Columns (2)–(5) of **Table 6** begin this analysis by examining the effect of the reform by broad offence categories, with the baseline set of controls. Note that though these categories are not mutually exclusive – individuals can be charged with multiple offences in different broad offence categories – in practice, there is not much overlap across categories. For instance, just 4% of sex offence cases with a seated jury include an additional charged classified as violent and less than 1% include a property or other offence. The full set of controls, including victim and defendant characteristics and proxies for the

<sup>36</sup> We only look at changes in sentence length for plea cases (as opposed to jury trials), since:

- (i) in theory, judges could respond to changes in jury acquittal behaviour with changes in sentencing behaviour and
- (ii) a change in jury acquittal behaviour would translate to a change in sentencing if treating acquittals as sentences of zero months.

<sup>37</sup> Appendix **Table B3** demonstrates that these results (or lack thereof) are seen in both our baseline three-year window and when defining the window as two or four years, as well as in both courts. These results are also not sensitive to including month, defendant and victim characteristics, number of charges and proxies for quality of evidence and severity of case – trial length, trial breaks and undefended.



Table 6  
*Is There a Post-reform Effect on Conviction by Broad Offence Category?*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Dependent variable = guilty_anyoff									
	All cases	Any violent	Any property	Any sex	Any other	All cases	Any violent	Any property	Any sex	Any other
Post-reform three years	0.001 (0.025)	-0.066 (0.042)	-0.047 (0.0471)	0.175 <sup>***</sup> (0.0629)	-0.044 (0.0501)	-0.0194 (0.025)	-0.097 <sup>**</sup> (0.044)	-0.126 <sup>***</sup> (0.047)	0.161 <sup>**</sup> (0.067)	-0.047 (0.053)
_cons	0.592 <sup>***</sup>	0.620 <sup>***</sup>	0.469 <sup>***</sup>	0.284 <sup>**</sup>	0.759 <sup>***</sup>	0.428 <sup>***</sup>	0.339 <sup>***</sup>	0.200 <sup>*</sup>	0.419 <sup>*</sup>	0.656 <sup>***</sup>
N	2,027	724	555	289	516	1,911	658	534	285	489
R <sup>2</sup>	0.049	0.029	0.104	0.141	0.054	0.111	0.134	0.244	0.267	0.132
Mean dependent variable	0.60	0.67	0.59	0.46	0.62	0.60	0.67	0.59	0.46	0.62
Court dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Offence types	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case characteristics 1	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Month dummies	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Case characteristics 2	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes

*Notes.* Robust standard errors in parentheses; \*10%, \*\*5%, \*\*\*1%. Post-reform three years is a dummy equal to one for the first three years after the reform: 1921, 1922 and 1923. All specifications also control for a dummy after the treatment period: 1924-6. Other case characteristics 1 include victim characteristics (male, female, and child), defendant gender, number of defendants, co-defendant id, no. charges and charged with criminal history. Other case characteristics 2 include trial length, undefended and jury breaks.

Table 7  
*Robustness of Sex Offence Results*

	(1)	(2)	(3)	(4)	(5)
	Dependent variable:				
	guilty_anyoff	guilty_shareoff	guilty_alloff	guilty_1stoffs	guilty_anyoff
	<i>Restricted to the sample with at least one sex offence</i>				
Post-reform three years	0.175*** (0.063)	0.169*** (0.060)	0.155** (0.061)	0.173*** (0.062)	
Year = 1921					0.061 (0.080)
Year = 1922					0.191** (0.092)
Year = 1923					0.333*** (0.091)
_cons	0.284** (0.125)	0.338*** (0.125)	0.453*** (0.141)	0.403*** (0.147)	0.295** (0.127)
N	289	289	289	285	289
R <sup>2</sup>	0.141	0.136	0.135	0.122	0.159
Post-reform two years	0.116* (0.068)	0.107 (0.065)	0.103 (0.067)	0.0998 (0.067)	
Post-reform four years	0.129** (0.059)	0.124** (0.056)	0.112* (0.057)	0.127** (0.058)	
Post-reform five years	0.141** (0.059)	0.134** (0.056)	0.118** (0.057)	0.140** (0.057)	

Notes. Robust standard errors in parentheses. \*10%, \*\*5%, \*\*\*1%. The upper panel presents the baseline treatment period of three years, and controls for all years after that. The lower panels redefine the treatment period to be two years (1921–2), four years (1921–4) or five years (1921–5) and include a dummy capturing all additional years (post-treatment period). All specifications control for court and the basic set of offence characteristics.

quality of evidence and case severity, are included in columns (6)–(10). While there is no effect of the reform on overall conviction rates, there is evidence that the reform had a significant positive impact on the likelihood of conviction for sex offence cases, as shown in columns (4) and (9).<sup>38</sup> We find that sex offence cases tried after the reform are about 17 percentage points more likely to result in conviction; given that just 46% of such cases result in conviction, this reflects a 37% increase in the likelihood of conviction following the reform.

Table 7 demonstrates the robustness of the sex offence findings to alternative dependent variables (the share of charges resulting in conviction, conviction of all charges and conviction on the first, most serious charge) and post-reform treatment windows (two, four and five years). The effect decreases in size with alternative windows, but is still quite large: depending on the window, exposure to the reform increases the chance of conviction of a sex offence by 12–17 percentage points. In additional specifications not shown, this robust finding for sex offences is still seen (at standard significance levels

<sup>38</sup> 43% of sex offence cases include at least one charge that is classified as a rape (or indecent assault) while the remainder is classified as ‘other sex offence’, which is largely composed of gross indecency. Sex offence defendants are predominantly male (97%). The majority (53%) of sex offences are classified as having no victim, while 5% have male victims and 42% have female victims.

and despite reduced sample sizes) when looking separately at the First *versus* Second Courts, and rape *versus* other sex offence cases (composed primarily of gross indecency).

Is it really the female jury reform that is underlying the sex offence effects? As the analysis relies on a simple pre-post design, we could potentially find the same effects, even if the reform has no impact, if society in general is becoming more and more sympathetic to female rights. Clearly, at this time, there was an increased acceptance of the female rights movement, in particular with respect to suffrage and females autonomy over their bodies. For instance, females over age 30 (with property qualifications) were given the right to vote in 1918; this was extended to women over 21 in 1928. In addition, the first permanent birth control clinic, the Mothers Clinic, opened in North London in 1921, which advised mothers on birth control.<sup>39</sup>

To assess the extent to which other societal changes may explain our results, we examine the pre-reform trends (to the extent feasible in this short time span) in conviction rates for sex offences in the top panel of Figure 2. We see no substantive upward trend in sex offence convictions prior to the reform; in fact, sex offence conviction rates are quite flat. Figure 2 also highlights the fact that the increase in conviction rates did not occur until 1922. The one year lag in the effect for sex offences actually supports the story that the effect is driven by seated female jurors and not changing attitudes in society (or just a handful of special cases for which females were seated), since less than 20% of sex offence cases had a female juror in 1921 compared to almost 60% in 1922 and 100% in 1923. Column (5) of Table 7 presents the regression equivalent, looking at the effect of the reform in each of the three treatment years separately; the effect of the reform on sex offence convictions in the first year after the reform (when hardly any females were seated) is an order of magnitude smaller than in 1922 and 1923 and insignificant. Finally, it is worth noting that less than half of the sex offences can be generalised as violent offences towards women, which we categorise as a rape; the remainder of other sex offences is generally classified as victimless. Yet, as mentioned above, the effect of the reform is seen for both types of sex offences.

In contrast to the increased post-reform conviction rate for sex offences, Table 6 shows there is a reduction in the chance of conviction for violent offences, property offences and other offences. Though these relationships are not significant when just controlling for the vector of offence type, the negative coefficients persist with the full set of controls and become significant for violent offences and property offences. These estimates are sizable in magnitude and imply a reduction in the likelihood of conviction of 10 and 13 percentage points, respectively. However, our analysis (not shown) of the robustness of the property offence findings do not tell as robust a story as that for sex offences – that is, the reduction in the conviction rate for property offences appears to be driven by the third year after the reform and is only visible with the full set of controls. Given the sensitivity of this finding and the fact that property offences are in fact the least female salient offence category (in the sense that it does not have a relatively high share of

<sup>39</sup> We thank a referee for pointing out this important and relevant event.

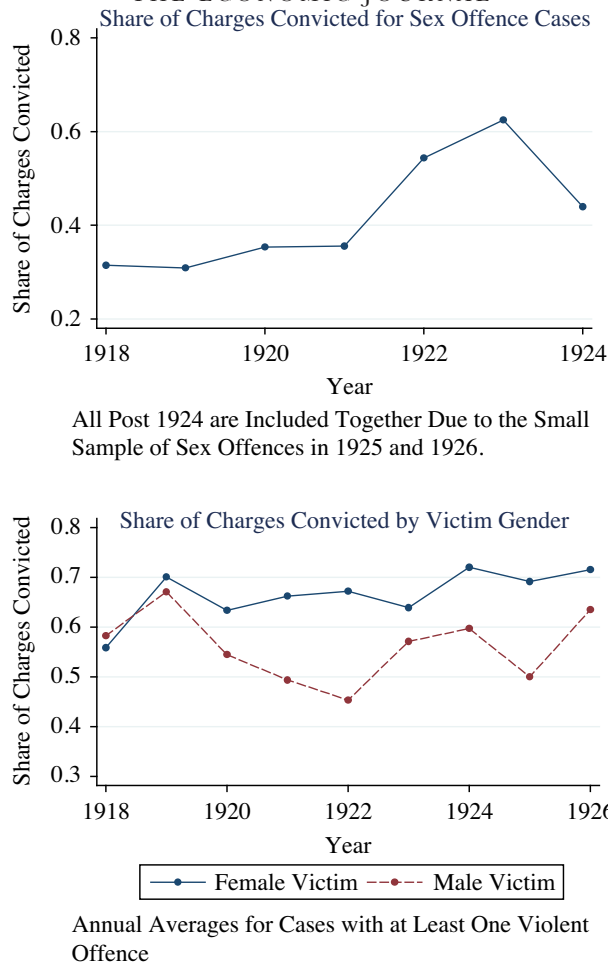


Fig. 2. *Could Underlying Trends in Attitudes Towards Females Explain the Main Results?*

Note. Colour figure can be viewed at <https://academic.oup.com/ectj>.

female victims or female defendants), we do not take forward the post-reform reduction in property convictions as a main result.

To begin to get at the potential importance of the female saliency of the case, we look directly at whether there was a differential effect of the reform on convictions for cases with at least one female victim. This would be in line with previous work that has shown that the impact of jurors can depend on victim gender (Anwar *et al.*, 2015). We present this analysis in Table 8; importantly, we restrict the sample to violent offences, as the classification of victim gender is the most straightforward for these cases: almost all violent offences have a male or female victim, whereas the majority of sex, property and other offences are classified as having no victim. Note that violent offences do not include cases of sexual violence, like rape, which are categorised as sex offences. We present the results of estimating (1) with a three-year post-reform window, and include an interaction of whether there is a female victim with all

Table 8  
*Heterogeneous Effect on Female Victim-violent Crime Cases*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Dependent variable:							
	guilty_any offence		guilty_shareoff	guilty_alloff		guilty_1stoff		
	Y = 3	Y = 3	Y = 3	Y = 3	Y = 3	Y = 2	Y = 4	Y = 5
Post-reform 'Y' years	-0.108 <sup>*</sup> (0.058)	-0.174 <sup>***</sup> (0.056)	-0.152 <sup>***</sup> (0.052)	-0.153 <sup>***</sup> (0.055)	-0.156 <sup>***</sup> (0.056)	-0.211 <sup>***</sup> (0.060)	-0.198 <sup>***</sup> (0.054)	-0.123 <sup>***</sup> (0.052)
Post-reform 'Y' years × female victim	0.146 <sup>*</sup> (0.088)	0.200 <sup>***</sup> (0.087)	0.191 <sup>**</sup> (0.083)	0.207 <sup>***</sup> (0.089)	0.208 <sup>***</sup> (0.090)	0.256 <sup>***</sup> (0.098)	0.162 <sup>*</sup> (0.084)	0.161 <sup>***</sup> (0.082)
female_victim	0.058 (0.067)	-0.0004 (0.068)	-0.010 (0.065)	-0.041 (0.069)	-0.021 (0.070)	-0.003 (0.068)	0.004 (0.068)	0.002 (0.068)
_cons	0.579 <sup>***</sup> (0.085)	0.372 <sup>***</sup> (0.10)	0.501 <sup>***</sup> (0.093)	0.644 <sup>***</sup> (0.100)	0.618 <sup>***</sup> (0.105)	0.404 <sup>***</sup> (0.101)	0.409 <sup>***</sup> (0.101)	0.409 <sup>***</sup> (0.100)
Mean dependent variable	0.67	0.67	0.59	0.51	0.54	0.67	0.67	0.67
Court dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Offence categories	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case characteristics 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month dummies	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case characteristics 2	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	661	658	658	658	644	658	658	658
R <sup>2</sup>	0.069	0.141	0.159	0.232	0.205	0.143	0.132	0.135

*Notes.* Robust standard errors in parentheses, \*10%, \*\*5%, \*\*\*1%. Post-reform three years is a dummy equal to one for the first three years after the reform (1921, 1922 and 1923). All specifications also control for a dummy after the treatment period: 1924–6, and an interaction of that post-treatment period with female victims. Columns (6)–(8) define the post-reform treatment window as two, four and five years, respectively. Case characteristics 1 include victim characteristics (male, female, no. and child), defendant gender, number of defendants, codefendant id, no. charges and charged with criminal history. Case characteristics 2 include trial length, undefended and jury breaks.

post-reform dummies. The first five columns present the three-year window results for alternative dependent variables and controls.<sup>40</sup> Columns (6)–(8) demonstrate robustness to alternative post-reform windows. Regardless of the specification, the impact of the reform on the conviction rate differential between male and female victim cases is significant and large in magnitude. The specification in column (2) implies that prior to the reform there was essentially no difference in the conviction rates between violent crime cases with female victims and those with male victims. Post reform, this conviction rate differential dramatically increased such that cases with female victims were now 20 percentage points more likely to be convicted than those with male victims. Thus, the reform had a significant impact on the way male *versus* female victim cases were treated.<sup>41</sup>

Note that the increase in the conviction differential between male and female victim is primarily driven by a decrease in conviction rates for male victim cases following the reform. This explains why the overall impact on violent crime conviction rates (regardless of victim gender) shown in Table 5 is negative. This can be seen both in the regression results and in the lower panel of Figure 2, which presents the average annual share of charges convicted for cases with any violent offences, broken down by victim gender. While perhaps a bit counterintuitive, a possible explanation for this combination of results is that female jurors have both a lower propensity to convict defendants in violent crime cases in general and a comparatively higher rate of convicting defendants in cases with a female *versus* male victim. Figure 2 also suggests that annual violent offence conviction rates are not trending one way or the other before the reform and that, prior to the reform, conviction rates for violent offences with male and female victims are moving together. Of course, however, the short pre-reform time window does limit our ability to say anything concrete about trends.

Table 9 considers whether adding females to the jury pool differentially impacts the conviction rates in female defendant cases. Specifically, we present the results of estimating (1) with a three-year post-reform window, and include an interaction of whether there is a female defendant with all post-reform dummies. When considering all types of offences in columns (1) and (2), we find that the post-reform period is negatively associated with the chance of any conviction for female defendants. The point estimates are large but imprecisely estimated, which is perhaps not surprising given that less than 13% of the jury trials are for female defendant cases. Thus, the remaining columns of Table 9 focus on the two offence categories with the greatest share of female defendants – violent offences (16%) and other offences (20%). While there is no evidence of differential treatment of female defendants in violent offence cases, the results

<sup>40</sup> All specifications control, however, for offence types and basic offence characteristics; this is important given the variation in offence type across violent offences with male *versus* female victims. A substantially higher share of female victim violent offences are for killing than for male victims (75% *versus* 46%) while males are more likely to be a victim of a robbery (18% *versus* 7% of cases). The point estimates are also fairly robust, albeit a little less precise, to including interactions between the full set of controls and the post-reform dummy; it is also worth noting that these additional interactions are generally not significant and small relative to the female victim interaction.

<sup>41</sup> Similar results (not shown) are also seen when looking separately by court. In addition, these results are driven by violent offences with male defendants; there are simply not enough female defendants – violent crime cases to conclude anything substantial (there are only 90 trials over the entire period).





Table 10  
Effect of the Reform on Alternative Verdicts and Hung/Dismissed Juries

	(1)	(2)	(3)	(4)	(5)	(6)
	Guilty of a lesser offence		Jury dismissed any reason	Hung jury	Jury discharged on at least one charge	Jury dismissed due to defendant confession
<i>Panel (a): include court and basic set of offence characteristics</i>						
Post-reform three years	−0.020** (0.008)	−0.036*** (0.013)	0.027** (0.012)	−0.003 (0.007)	0.023*** (0.006)	0.011 (0.008)
Mean dependent variable	0.025	0.04	0.083	0.025	0.029	0.031
N	2,027	1,225	2,148	2,148	2,148	2,148
R <sup>2</sup>	0.058	0.082	0.028	0.015	0.063	0.022
<i>Panel (b): also includes extended set of case characteristics</i>						
Post-reform three years	−0.015** (0.008)	−0.028** (0.013)	0.023* (0.013)	−0.002 (0.008)	0.019*** (0.006)	0.008 (0.009)
Mean dependent variable	0.022	0.036	0.085	0.026	0.029	0.032
N	1,920	1,154	2,037	2,037	2,037	2,037
R <sup>2</sup>	0.051	0.073	0.042	0.024	0.079	0.033
Sample	All jury verdicts	Guilty jury verdicts	All jury seated	All jury seated	All jury seated	All jury seated

Notes. Robust standard errors in parentheses. \*p<0.10, \*\*p<0.05, \*\*\*p<0.01. All specifications also control for whether the case is after the main post-reform period.

presented in Columns (5)–(9) are strongly suggestive that adding females to the jury pool decreases conviction rates for female defendants charged with ‘other’ offences. Adding females to the pool decreases the chance of conviction of female defendants charged with ‘other’ offences by about 18 percentage points. Though not quite significant using the three-year window in our baseline specification, the relationship is significant with two-year and four-year post-reform windows. Digging into the offences that comprise the ‘other’ category helps to understand the source of the effect: at least one-third of female defendants charged with offences in this category are charged with obviously female ‘salient’ offences, including abortion, aiding abortion, conspiracy to commit abortion and procurement of a miscarriage.

Finally, Table 10 considers whether there is an effect on outcomes other than convictions. Panel (a) includes just the basic set of offence types, while panel (b) includes the extended set of case characteristics. The dependent variable in columns (1) and (2) is whether the jury finds the defendant guilty of a lesser charge; i.e. they are convicted of an offence of lesser severity than the original charge. Column (1) uses our baseline analysis sample of all jury verdicts and finds that there is a 1.5–2 percentage point reduction in the chance of a conviction of a lesser offence. In this sample, however, such a reduction can either be driven by an increase in the chance of being acquitted or the chance of being convicted of the original charge. To disentangle this, column (2) restricts the analysis to guilty defendants and still finds a significant reduction in the chance of conviction a lesser charge (of around 3 percentage points), pointing to this relationship being driven by an increase in conviction of the full charge. Finally, conducting this analysis at the broad offence category indicates that these results are primarily

driven by violent offences, whereby for instance, a charge of murder could be reduced to manslaughter.

The remaining columns of Table 10 examine the sample of all cases for which a jury was seated, regardless of whether the jury reached a verdict. Column (3) tests whether there is a post-reform effect on the likelihood of having a jury dismissed for any reason including a hung jury, being discharged on at least one charge when typically convicted of other charges, and being dismissed on select charges when the defendant pleads to other charges after jury selection. Columns (4)–(6) look at the specific types of dismissals separately. Column (3) demonstrates that, overall, there is an increase in the likelihood of a jury being dismissed for any reason by 2.7 percentage points; given that only 8% of seated juries are dismissed on at least one charge, this is a substantial increase.<sup>42</sup> Most of the overall effect appears to be driven by juries being discharged on at least one charge (while typically reaching a verdict (usually guilty) on other charges). When looking at offence subsamples (not shown), a significant effect is seen for property offences, sex offences and other offences, while the point estimate for violent offences is large but not quite significant. In contrast, there is no significant effect on the likelihood of a hung jury or the jury being dismissed due to a defendant confession.<sup>43,44</sup>

One possible explanation is that having reached a verdict on some charges, the other charges are dismissed to avoid a lengthy deliberation. Consistent with such an explanation, we find that trials became significantly longer after the reform. As seen in Appendix Table B4, trials increased in length (including deliberation) by approximately 0.15 to 0.3 days (or 8–15%, compared to the average length of 1.9 days), depending on the specification. This could have occurred, for instance, if females were more careful in evaluating evidence or less familiar with the laws, or increased the diversity of perspectives within juries.

#### 4.4. *The Effect of Seated Female Jurors*

As discussed in detail above, our focus on the pre-post reform analysis is driven by concerns that the pre-trial process of seating juries is not random and may lead, therefore, to spurious correlation between the gender composition of the jury and conviction rates. While avoiding issues related to non-random jury selection, a disadvantage of the pre-post design is that the identified effect includes both the direct effect of seating females on the jury and any indirect effects associated with how pre-trial selection is conducted in the presence of female jurors. Perhaps more importantly, the pre-post design also

<sup>42</sup> The effect is seen regardless of the post-reform window used, though it appears to get substantially larger four to five years after the reform (results available upon request). The overall effect in the three-year window appears to be driven by property offences and other offences, though as the window is expanded, the same effect is also seen for violent offences. Such an effect is never seen for sex offences.

<sup>43</sup> It should be noted, however, that the point estimate for sex offences is quite large (though imprecise); sex offence cases are more than three percentage points less likely to result in a hung jury after the reform.

<sup>44</sup> Given these findings, it is also not surprising that including these cases for which the jury seated but did not reach a verdict in the denominator does not affect the main results. Specifically, Appendix Tables B5 and B6 demonstrate the robustness of our results to treating a hung verdict as an acquittal, and assigning either guilty or not guilty to a discharged jury.

does not allow us to say anything about the effect of each additional seated female on the jury.

This Section seeks to look directly at the relationship between the number of seated female jurors and verdicts. To this end, we conduct three exercises. In the first two, we step away from our pre-post analysis and focus instead on just the post-reform period; specifically, we estimate regressions of the case outcome (e.g. share of charges convicted) on the number of females seated on the jury. The first exercise utilises a sub-sample of cases that were decided by 'repeat juries', i.e. juries that were allowed to sit for multiple trials without a new jury being drawn, while the second exercise focuses on jury trials with at least one seated female juror. Both of these analyses have the advantage of utilising case-by-case variation in the gender composition of the jury (due, for example, to day to day variation in the composition of the pool), rather than limiting the variation to that due to the pre-post change in the master jury books.

The primary concern with exploiting variation in the seated jury is that females might be seated on cases that are distinctly different from the cases where they are not seated (but are in the pool). Our first exercise takes two steps to limit such concerns and facilitate a causal interpretation of our results. First, we focus on the violent crimes–female victims findings, as female victims do not predict whether females are seated on the jury overall for the subsample of violent crimes (see Appendix Table B1); we purposely do not focus on crimes, such as sex offences, where there is much stronger evidence of gender differences in jury selection. Second, we restrict our sample to cases decided by a jury that faced multiple trials; note that we include all trials by that jury. Recall that a unique feature of this time period is the use of the same jury on multiple cases. We argue that there should be less selection on specific case characteristics when the same jury is carried over from a previous trial; this could occur because the second case was not known before the jury was chosen (though we do not have explicit evidence of this) or because the jury seated for the first case was not tailored to a particular kind of case, perhaps because the case was relatively less serious.<sup>45</sup> While we do not have any anecdotal evidence or rules informing us about when repeat juries can be used, an analysis of the data sheds some light on this. First, we find in regressions not shown that whether a case has a repeat jury is not random, as case characteristics are significantly related to repeat jury use: generally speaking, characteristics that may be correlated with less serious cases (e.g. Court 2 and female defendants) increase the chance of a repeat jury, while characteristics suggestive of more complicated or serious trials (e.g. multiple defendants, conspiracy charges, robbery) decrease the likelihood. We also note that most repeat juries occur over one day, though they occasionally spill over onto a second day.

Restricting the data to juries that decided multiple trials results in a sample of 235 violent crime trials (130 unique juries) with complete verdict and victim gender information; 14% of these cases have zero females seated, 20% have one female, 26% have two females, and the remainder has three or more. Note that the variation in the gender composition of the seated jury is across and not within juries, and hence jury fixed effects cannot be included (though we do cluster the standard errors on the jury).

<sup>45</sup> Indeed, regressions (available upon request) of the presence of female jurors at the extensive or intensive margin on the full set of case characteristics have R-squares that are 2–3 times as large when using the sample of one case juries compared to multiple case juries.

Rather, we are comparing verdicts for juries with relatively more females seated to juries with relatively fewer females, and arguing that the number of females seated for this sample of repeat juries is not driven by selection. Instead, juries with more seated females were likely picked on days when there was a jury pool with relatively more females.<sup>46</sup>

Results are presented in Table 11. Columns (1)–(3) of Table 11 examine the effect of the number of seated female jurors on various measures of conviction (any, share, 1st offence), and includes an interaction between the number of seated females and whether the victim is female. The results are quite robust across alternative dependent variables. Given the possibility of selection on the basis of case characteristics, we control for observable characteristics; however, the results are not sensitive to such controls, which is again suggestive of the number of seated females not being driven by case specific selection. We find that each additional female seated significantly increases the likelihood of conviction on female victim offences by more than 10 percentage points on average, whereas the number of seated females do not directly affect the conviction rate for male victim cases. These effects are quite large given that 60% of defendants are convicted of at least one charge. Columns (4)–(6) assess whether the observed relationship is indeed linear or whether there is a differential effect of adding two or more female jurors. Relative to seated juries with zero females, adding just one female to the jury increases the chance of conviction on female victim cases by slightly more than 10 percentage points; though this estimate is very imprecisely measured, it is in line with the estimates from the linear specifications. However, adding two or more females to the seated jury significantly increases female victim conviction rates by 30–40 percentage points. These results suggest that it is not just the presence of any females that affect the verdicts, but that there is added influence of additional female voices during deliberations.<sup>47</sup>

Columns (7) and (8) present the results of our second exercise, which restricts the sample to 352 post-reform violent crime trials where at least one female is seated. For this analysis, we are focusing on a sub-sample of violent crime cases that were not deemed ‘too sensitive’ for female ears, and implicitly assuming that the number of females seated (conditional on at least one female being seated) does not reflect case characteristics. These results once again indicate that each additional female juror seated (column (7)) and that having two or more females seated compared to just one (column (8)) increases the conviction rate for female victim cases. While these results are for a selected sample of cases, they do highlight the fact that the findings are not driven by just having at least one female on the jury.<sup>48</sup>

Finally, in our third exercise presented in Table 12, we return to the full sample (before and after the reform) and use an instrumental variable approach. Specifically, we

<sup>46</sup> It should be noted that restricting the analysis to violent crime repeat juries is potentially selecting on a subset of violent crimes; this can be seen in the conviction rates; the conviction rate for repeat jury violent crimes is 61% throughout the sample period compared to 73% for single use jury violent crimes.

<sup>47</sup> Specifications (not shown) that look separately at one, two and three or more seated females find similar patterns; larger point estimates on the interactions for two *versus* one and three *versus* two seated females.

<sup>48</sup> These results are also robust to how the dependent variable is defined and including or excluding controls. Finally, conducting the same exercise of looking at juries with at least one seated females for the sample of other offences finds the same pattern of results for female defendants as in the before–after analysis; however, they are generally only significant at around the 15% level.

Table 11  
*The Direct Effect of Seated Female Jurors on Violent Crime Convictions*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Share of guilty offences	Guilty of any offence	Guilty of 1st offence	Share of guilty offences	Guilty of any offence	Guilty of 1st offence	Share of guilty offences	Share of guilty offences
No. female jurors	−0.004 (0.027)	−0.010 (0.030)	0.001 (0.028)				−0.022 (0.024)	
No. female jurors × female victim	0.114** (0.046)	0.102* (0.052)	0.119** (0.047)				0.076* (−0.045)	
One female juror				0.038 (0.158)	0.004 (0.170)	0.050 (0.167)		
One female juror × female victim				0.062 (0.230)	0.117 (0.250)	0.050 (0.230)		
Two or more female jurors				0.029 (0.122)	0.006 (0.132)	0.055 (0.128)		−0.051 (0.075)
Two or more female jurors × female victim				0.370** (0.185)	0.333* (0.199)	0.398** (0.186)		0.281** (0.118)
Female victim	−0.175 (0.122)	−0.153 (0.134)	−0.193 (0.124)	−0.188 (0.171)	−0.179 (0.186)	−0.211 (0.169)	−0.011 (0.122)	−0.047 (0.106)
_cons	0.651*** (0.151)	0.476*** (0.161)	0.847*** (0.155)	0.628*** (0.160)	0.462*** (0.174)	0.812*** (0.164)	0.732*** (0.139)	0.723*** (0.145)
Offence categories and court	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other offence characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Just repeat juries?	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Condition on at least one seated female	No	No	No	No	No	No	Yes	Yes
N	235	235	231	235	235	231	352	352
R <sup>2</sup>	0.174	0.167	0.220	0.182	0.169	0.234	0.097	0.106

*Notes.* Robust standard errors, \*10%, \*\*5%, \*\*\*1% (clustered on jury in columns (1)–(6)). This analysis is restricted to cases with at least one violent offence in the post-reform period, as well as those trials in which the jury was a repeat jury, i.e. sat on more than one trial, in columns (1)–(6), and juries with at least one seated female in columns (7) and (8). This implies that in column (8) the omitted category is juries with at least one female (as opposed to zero females in columns (4)–(6)). Other offence characteristics include defendant gender, number of defendants, defendant order, victim child, criminal history and number of charges.

Table 12  
*The Direct Effect of Seated Female Jurors – Instrumental Variable Analysis*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Dependent variable = share guilty off						
	All	Prop	Sex	Other	Violent	Viol, femvic	Viol, nofem
Panel (a):							
any_female_jurors	0.020 (0.035)	−0.093* (0.055)	0.517*** (0.198)	−0.012 (0.069)	−0.033 (0.054)	0.061 (0.091)	−0.084 (0.063)
F-statistics	927	602	41	344	405	132	380
N	1,899	537	285	481	655	245	410
R <sup>2</sup>	0.074	0.174	−0.042	0.081	0.070	0.106	0.081
Panel (b):							
num_female_jurors	0.007 (0.013)	−0.031* (0.019)	0.195*** (0.074)	−0.004 (0.026)	−0.013 (0.022)	0.026 (0.038)	−0.033 (0.025)
F-statistics	435	241	38	147	215	92	159
N	1,899	537	285	481	655	245	410
R <sup>2</sup>	0.075	0.165	−0.091	0.080	0.071	0.106	0.084
Court dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Off type controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other char	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes. Robust standard errors in parentheses, clustered by jury. This table looks at the effect of having at least one seated female juror (panel (a)) or the number of seated jurors (panel (b)) on the share of charges convicted overall and by offence. Whether any females are seated or the number of females seated is instrumented for with a leave out share of females seated on other juries in the same month and court; this can be seen as a proxy for the female composition of the jury pool, which goes from 0 to positive after the reform and then has some month to month, and annual to annual variation (as the jury books are replaced). First stage F-statistics are reported for all subsamples.

regress the share of guilty verdicts on whether there is at least one female seated on the jury (panel (a)) and the number of females seated on the jury (panel (b)), instrumenting for the gender composition of the seated jury with the share of females seated on other juries in the *same* court and month, making sure to exclude the current jury from the calculation. Since we unfortunately do not have data on the actual pool from which the seated jury was chosen (i.e. the ideal instrument), we proxy for the gender composition of the pool with that from other cases seen in the same court and time period. Much of the variation in this instrument still comes, however, from the addition of females to the pool after the reform, though there can still be some year to year variation as the master books are updated annually or month to month variation if there are changes in the composition of who responds to jury summons over time. This instrument must be created separately for each court, as the pools are drawn from different master books; therefore, the composition of the seated jury in the second court provides little information about the pools in the first court.<sup>49</sup>

This exercise yields the same basic results as our main pre-post reduced form analysis: the addition of females to the seated jury significantly increases the chance of

<sup>49</sup> The first stage F-statistics are well above weak instrument thresholds and noted in the table; this is seen overall and across sub-samples.



conviction for sex offences and decreases (marginally significant) the chance of conviction for property offences. When breaking violent crimes into female and non-female victim cases (columns (6) and (7)), we find that adding at least one female to the seated jury increases the chance of conviction for female victim cases by six percentage points but decreases the chance of conviction for non-female conviction cases by seven percentage points; while this gap is consistent with our earlier results, it is not precisely estimated in this specification.

## 5. Conclusion

This article takes advantage of a novel data set and a historical jury reform to study the effect of female jurors on real-world jury verdicts – a difficult question to study today given that seated juries are not random and that jury pools and seated juries are typically about 50% female. Though there was no overall effect of the reform on conviction rates, our findings suggest that adding females to the jury pool significantly impacted conviction rates on cases that may be particularly salient for females – namely sex offences, violent offences with female *versus* male victims and female defendants charged with highly female salient (abortion related) offences. The magnitudes of these results are substantial, especially considering that an average of two females serve on each 12-person jury in the post-reform period.

The findings of the pre-post reform analysis can be driven by both direct and indirect mechanisms. With respect to the former, complementary regression analyses suggest that the seating of female jurors increased the conviction rate differential between violent offence cases with female *versus* male victims. Given the strong evidence that females are (de)selected from sex offence cases, especially in the early years, we are limited in the extent to which we can look at the direct channel in this instance. On the other hand, the fact that the sex offence effect did not occur until one year after the reform – when females began to serve on such cases in substantially larger numbers – is suggestive again of the direct channel of seated female jurors.

Taken together, the results of the pre-post reform and seated jury regression analyses paint a very consistent picture, implying that the addition of female jurors substantially increased the conviction differential for violent crime cases between female *versus* male victim cases. One potential explanation is that all else equal (i.e. holding the case characteristics and quality of evidence constant), all-male juries prior to the reform did not deem the alleged violence against women to be illegal in a subset of cases (e.g. domestic violence). An alternative explanation is that female jurors are not impartial, perhaps because they overly sympathise with the female victim. Of course, these findings could also be generated by mechanisms that are more benign in nature. For instance, the quality of evidence may not be constant across cases if female jurors are simply better at assessing the testimony of a female victim than a male victim. In general, without knowing more about the exact nature of the evidence, it is impossible to distinguish among these explanations. An important consideration in weighing these possibilities, however, is that juries were still overwhelmingly male after the reform, requiring female jurors to affect the votes of a large number of male colleagues in order to impact trial outcomes.

While the nature of any historical study raises obvious issues about the generalisability to modern times, it is worth setting these concerns alongside the advantages of studying such an important moment in history. In the context of examining the impact of female jurors, these advantages include the ability to deal with the serious endogeneity issues related to the pre-trial selection process and to study the broad effects of female representation as opposed to variation around balanced gender composition. Our study also contributes to the much larger literature on the effects of female representation in political, civic, social and economic life. In this broader context, the study of juries avoids a fundamental selection bias that hampers efforts to reach causal conclusions in the vast majority of these settings. In particular, elected or appointed individuals must naturally satisfy the preferences of the underlying constituency or individual making the appointment, as is the case for judicial appointments, elected officials or positions in corporate leadership. This ubiquitous selection problem generally tends to diminish gender differences among selected individuals, making it exceedingly difficult to identify causal effects of female representation in most aspects of political or civic life.

## Appendix A. Data Source Appendix

Original data for this project was obtained from the Central Criminal Court: Court Books, which are held at the National Archives in London. The following six books were used:

Central Criminal Court: Court Books, CRIM 6/25 October 1917–September 1920 First Court, National Archives.

Central Criminal Court: Court Books, CRIM 6/26 October 1920–July 1923 First Court, National Archives.

Central Criminal Court: Court Books, CRIM 6/27 September 1923–October 1926 First Court, National Archives.

Central Criminal Court: Court Books, CRIM 6/63 January 1918–September 1920 Second Court, National Archives.

Central Criminal Court: Court Books, CRIM 6/64 October 1920–December 1922 Second Court, National Archives.

Central Criminal Court: Court Books, CRIM 6/65 January 1923–January 1925 Second Court, National Archives.

A photocopy of one page of the data source is provided in [Figure A1](#).

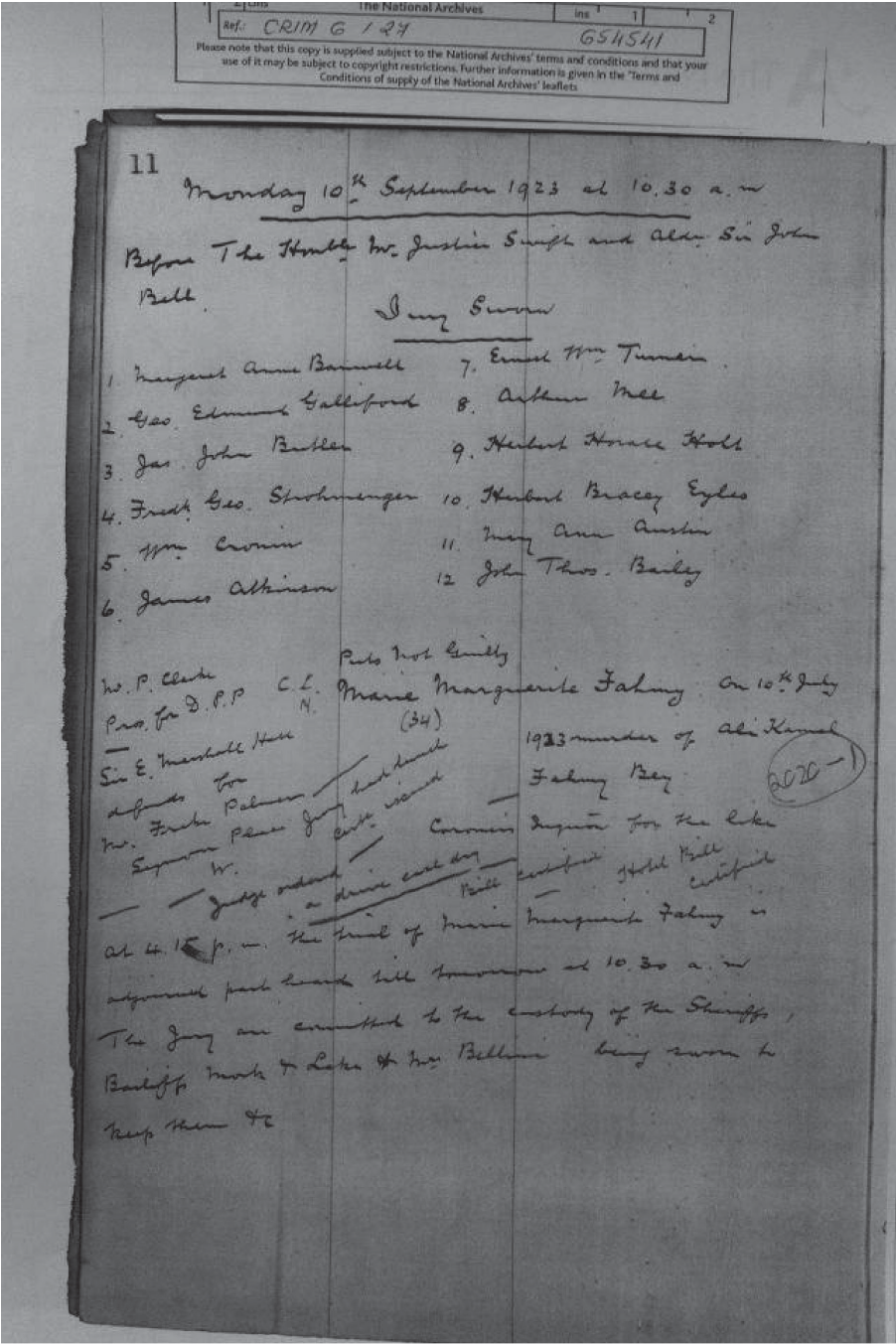


Fig. A1. Photograph of Jury Trial Raw Data from the First Court

**Appendix B. Additional Tables**

Table B1

*Average Female Juror Variables by Offence Category after the Reform (1921–6)*

	<i>N</i>	Any female jurors	No. female jurors
<i>By offence type</i>			
any_killing	287	0.81	1.86
any_ag_gass	122	0.87	2.28
any_arson	31	0.97	2.52
any_robbery	64	0.81	2.05
Any other off. <i>versus</i> person	118	0.75	1.92
any_burglary	122	0.94	2.56
any_theft	176	0.87	2.54
any_other_property	173	0.89	2.39
any_othersex	99	0.27	0.71
any_rape	87	0.56	1.57
any_deception	230	0.79	2.03
any_off_ag_king	40	0.85	2.23
any_bigamy	39	0.97	3.05
any_other	94	0.79	1.88
<i>By victim gender</i>			
Male victim	519	0.84	2.28
Female victim	329	0.79	2.04
No victim	434	0.71	1.74
<i>By defendant gender</i>			
Male defendant	1,165	0.77	2.05
Female defendant	177	0.85	2.04

Table B2

*Are Female Jurors Selected Based on Case Characteristics?*

	(1)	(2)	(3)
	Dependent variable = any_female_jurors		
Years	1921–6	1921–3	1924–6
Offence types	All	All	All
Repeat juries only?	No	No	No
<i>N</i>	1,267	832	435
Court 2	0.00156 (0.0261)	0.0774** (0.0354)	–0.100*** (0.0385)
any_killing	–0.0586 (0.0396)	–0.0403 (0.0531)	0.0590 (0.0517)
any_agg_assault	0.0652* (0.0389)	0.0104 (0.0561)	0.179*** (0.0370)
any_other_offence_against_person	–0.0610 (0.0481)	–0.0196 (0.0613)	–0.00408 (0.0645)
any_robbery	–0.0352 (0.0561)	–0.101 (0.0738)	0.182*** (0.0512)
any_rape	–0.234*** (0.0791)	–0.296*** (0.0862)	–0.0297 (0.159)
any_other_sex	–0.489*** (0.0595)	–0.495*** (0.0715)	–0.375*** (0.101)
any_arson	0.145** (0.0637)	0.168** (0.0756)	0.173 (0.121)

Table B2  
(Continued)

	(1)	(2)	(3)
	Dependent variable = any_female_jurors		
Years	1921–6	1921–3	1924–6
any_burglary	0.0980** (0.0410)	0.0496 (0.0534)	0.201*** (0.0629)
any_theft	0.0452 (0.0361)	0.00709 (0.0437)	0.201*** (0.0662)
any_other_property	0.0123 (0.0359)	0.0516 (0.0481)	–0.00964 (0.0630)
any_off_against_king	0.0480 (0.0618)	0.000762 (0.0845)	0.209*** (0.0490)
female_victim	0.00902 (0.0311)	0.0109 (0.0414)	–0.00215 (0.0458)
no_victim	–0.0519* (0.0296)	–0.0104 (0.0398)	–0.13*** (0.0431)
child_victim	0.0670 (0.101)	0.0690 (0.120)	0.0882 (0.175)
Defendant female	0.0350 (0.0354)	0.0640 (0.0457)	–0.0219 (0.0580)
No. counts	–0.00370 (0.00291)	–0.007** (0.00302)	–0.00021 (0.00181)
_cons	0.845*** (0.0455)	0.778*** (0.0605)	0.780*** (0.0640)
R <sup>2</sup>	0.165	0.170	0.276

Notes. Robust standard errors, \*10%, \*\*5%, \*\*\*1%. Not shown coefficients number of defendants, codefendant id, bigamy, conspiracy, deception and charged with criminal history.

Table B3  
*Post-reform Effect on the Likelihood of Conviction, Overall and by Court*

	(1)	(2)	(3)
	Dependent variable = guilty any offence		
	All	Court 1	Court 2
<i>Panel (a): three year window</i>			
Post-reform three years	0.00123 (0.0245)	–0.00811 (0.0352)	–0.00907 (0.0338)
_cons	0.592*** (0.0359)	0.694*** (0.0499)	0.374*** (0.0491)
N	2,027	998	1,029
R <sup>2</sup>	0.049	0.025	0.076
<i>Panel (b): two year window</i>			
Post-reform two years	–0.00738 (0.0270)	–0.0130 (0.0393)	–0.0185 (0.0368)
<i>Panel (c): four year window</i>			
Post-reform four years	0.0265 (0.0229)	–0.00337 (0.0333)	0.0283 (0.0319)
Court dummy and offence types	Yes	Yes	Yes

Notes. Robust standard errors in parentheses; \*10%, \*\*5%, \*\*\*1%. In panel (a), Post-reform three years is a dummy equal to one for the first three years after the reform: 1921, 1922 and 1923. All panel (a) specifications also control for a dummy after the treatment period: 1924–6. In panels (b) and (c), the treatment period is 1921–2 and 1921–4, respectively. For the ease of presentation, just the post-reform treatment effect is reported in panels (b) and (c).

Table B4  
*Is There a Post-reform Effect on Trial Length?*

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable = trial_length					
	three years		two years		four years	five years
Post-reform <i>Y</i> years	0.261*** (0.0865)	0.144* (0.0748)	0.164** (0.0791)	0.312*** (0.116)	0.159** (0.0745)	0.195*** (0.0724)
_cons	1.893*** (0.147)	0.420* (0.215)	0.404* (0.241)	1.871*** (0.146)	1.838*** (0.147)	1.873*** (0.149)
Mean dependent variable	1.9 days	1.9 days	1.9 days	1.9 days	1.9 days	1.9 days
Court and basic offence	Yes	Yes	Yes	Yes	Yes	Yes
Other characteristics 1	No	Yes	Yes	No	No	No
Month and other characteristics 2	No	No	Yes	No	No	No
<i>N</i>	2,143	2,032	2,032	2,143	2,143	2,143
<i>R</i> <sup>2</sup>	0.289	0.474	0.485	0.290	0.288	0.288

*Notes.* Robust standard errors in parentheses. \*10%, \*\*%, \*\*\*1%. All specifications control for whether the case is after the main post-reform period.

Table B5  
*Sensitivity of Sex Offence Results to Refining the Dependent Variable*

	(1)	(2)	(3)	(4)	(5)
	Dependent variable:				
	Baseline: guilty of any offence	Guilty of any, including pleas as guilty	Guilty of any, including hung as acquittal	Guilty of any, assigning acquit to discharged juries	Guilty of any, assigning convict to discharged juries
Post-reform three years	0.175***	0.161***	0.179***	0.146**	0.155**
Constant	(0.0629) 0.284** (0.125)	(0.055) 0.451*** (0.116)	(0.0615) 0.311** (0.123)	(0.0613) 0.292** (0.125)	(0.0628) 0.285** (0.121)
<i>N</i>	289	348	301	302	302
<i>R</i> <sup>2</sup>	0.141	0.222	0.138	0.131	0.124

*Notes.* Robust standard errors in parentheses. \*10%, \*\*%, \*\*\*1%. This table presents sensitivity analyses of the baseline specification in column (1) of Table 7 to redefining the dependent variables to include cases with seated juries but no full verdict. All specifications control for court and the basic set of offence characteristics.

Table B6  
*Sensitivity of Female Victim Results to Refining the Dependent Variable*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Baseline: Guilty of any offence	Guilty of any, including pleas as guilty	Guilty of any, including hung as acquittal	Guilty of any, assigning acquit to discharged juries	Guilty of any, assigning convict to discharged juries	Baseline: Guilty of any offence	Guilty of any, including pleas as guilty	Guilty of any, including hung as acquittal	Guilty of any, assigning acquit to discharged juries	Guilty of any, assigning convict to discharged juries
Post-reform three years	-0.108* (0.0575)	-0.0737 (0.0533)	-0.107* (0.057)	-0.120** (0.0566)	-0.0981* (0.0552)	-0.174*** (0.0556)	-0.143*** (0.0552)	-0.169*** (0.0547)	-0.186*** (0.0536)	-0.147*** (0.0543)
Post-reform three years × female vic	0.146* (0.0878)	0.1 (0.0813)	0.136 (0.0874)	0.137 (0.0883)	0.137 (0.0838)	0.200** (0.0869)	0.146* (0.0857)	0.182** (0.0866)	0.191** (0.0866)	0.181** (0.0834)
_cons	0.579*** (0.0846)	0.691*** (0.0768)	0.571*** (0.0846)	0.618*** (0.0857)	0.585*** (0.0809)	0.372*** (0.1)	0.397*** (0.0999)	0.375*** (0.0981)	0.393*** (0.0992)	0.401*** (0.0963)
Court	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic offence	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Case characteristics 1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Month and case characteristics 2	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
N	661	741	676	701	701	658	672	673	698	698
R <sup>2</sup>	0.069	0.052	0.074	0.071	0.068	0.141	0.129	0.146	0.159	0.123

*Notes.* Robust standard errors in parentheses; \* 10%, \*\* 5%, \*\*\* 1%. Post-reform three years is a dummy equal to one for the first three years after the reform: 1921, 1922 and 1923. All specifications also control for a dummy after the treatment period: 1924–6, and an interaction of that post-treatment period with female victims. Case characteristics 1 include victim characteristics (male, female, no and child), defendant gender, number of defendants, codefendant id, no. charges, and charged with criminal history. Case characteristics 2 include trial length, undefended and jury breaks.



RAND Corporation  
 Duke University and NBER  
 University of Gothenburg and CEPR

Accepted: 18 September 2017

Additional Supporting Information may be found in the online version of this article:

## Data S1.

## References

- Andreoni, J. and Vesterlund, L. (2001). 'Which is the fair sex? Gender differences in altruism', *Quarterly Journal of Economics*, vol. 116(1), pp. 293–312.
- Anwar, S., Bayer, P. and Hjalmarsson, R. (2012). 'Jury discrimination in criminal trials', *Quarterly Journal of Economics*, vol. 127(2), pp. 1017–55.
- Anwar, S., Bayer, P. and Hjalmarsson, R. (2014). 'The role of age in jury selection and trial outcomes', *Journal of Law and Economics*, vol. 57(4), pp. 1001–30.
- Anwar, S., Bayer, P. and Hjalmarsson, R. (2015). 'Politics in the courtroom: political ideology and jury decision making', NBER Working Paper No. 21145.
- Anzia, S.F. and Berry, C.R. (2011). 'The Jackie (and Jill) Robinson effect: why do congresswomen outperform congressmen?', *American Journal of Political Science*, vol. 55(3), pp. 478–93.
- Ashenfelter, O., Eisenberg, T. and Schwab, S.J. (1995). 'Politics and the judiciary: the influence of judicial background on case outcomes', *Journal of Legal Studies*, vol. 24(2), pp. 257–81.
- Baldwin, J. and McConville, M. (1979). 'Trial by jury: some empirical evidence on contested criminal cases in England', *Law and Society Review*, vol. 13(4), pp. 861–90.
- Bennion, F. (1979). 'The Sex Disqualification (Removal) Act – 60 inglorious years', *New Law Journal*, vol. 129, pp. 1088–9.
- Bentley, D. (1998). *English Criminal Justice in the Nineteenth Century*, London: The Hambledon Press.
- Bridgeman, D. and Marlowe, D. (1979). 'Jury decision making: an empirical study based on actual felony trials', *Journal of Applied Psychology*, vol. 64(2), pp. 91–8.
- Campbell, K. and Mínguez-Vera, A. (2008). 'Gender diversity in the boardroom and firm financial performance', *Journal of Business Ethics*, vol. 83(3), pp. 435–51.
- Carter, D.A., D'Souza, F., Simkins, B.J. and Simpson, W.G. (2010). 'The gender and ethnic diversity of US boards and board committees and firm financial performance', *Corporate Governance: An International Review*, vol. 18(5), pp. 396–414.
- Cockburn, J.S. and Green, T. (1988). *Twelve Good Men and True: The Criminal Trial Jury in England, 1200–1800*, Princeton, NJ: Princeton University Press.
- Cowan, C.L., Thompson, W.C. and Ellsworth, P.C. (1984). 'The effects of death qualification on jurors' predisposition to convict and on the quality of deliberation', *Law and Human Behavior*, vol. 8(1), pp. 53–79.
- Department of Defense (2014). 'Department of Defense annual report on sexual assault in the military, fiscal year 2013', available at: <https://www.army.mil/e2/c/downloads/343174.pdf> (last accessed: 26 September 2017).
- Devine, D., Clayton, L., Dunford, B., Seying, R. and Pryce, J. (2000). 'Jury decision making: 45 years of empirical research on deliberating groups', *Psychology, Public Policy, and Law*, vol. 7(3), pp. 622–727.
- Fischer, G.J. (1997). 'Gender effects on individual verdicts and on mock jury verdicts in a simulated acquaintance rape trial', *Sex Roles*, vol. 36(7–8), pp. 491–501.
- Fitzgerald, R. and Ellsworth, P. (1984). 'Due process vs. crime control: death qualification and jury attitudes', *Law and Human Behavior*, vol. 8(1), pp. 31–51.
- Fowler, L. (2005). 'Gender and jury deliberations: the contributions of social science', *William and Mary Journal of Women and the Law*, vol. 12(1), pp. 1–48.
- Gabora, N.J., Spanos, N. and Joab, A. (1993). 'The effects of complainant age and expert psychological testimony in a simulated child sexual abuse trial', *Law and Human Behavior*, vol. 17(1), pp. 103–19.
- Garvey, S., Hannaford-Agor, P., Hans, V., Mott, N., Munsterman, G.T. and Wells, M. (2004). 'Juror first votes in criminal trials', *Journal of Empirical Legal Studies*, vol. 1(2), pp. 371–98.
- George, T. (2001). 'Court fixing', *Arizona Law Review*, vol. 43(1), pp. 9–62.
- Goodman, G.S., Tobey, A., Batterman-Faunce, J., Orcutt, H., Thomas, S., Shapiro, C. and Sachsenmaier, T. (1998). 'Face to face confrontation: effects of closed-circuit technology on children's eyewitness testimony and jurors' decisions', *Law and Human Behavior*, vol. 22(2), pp. 165–203.

- Hastie, R., Penrod, S.D. and Pennington, N. (1983). *Inside the Jury*, Cambridge, MA: Harvard University Press.
- Hegelsson, V. and Shaver, K. (1990). 'Presumption of innocence: congruence bias induced and overcome', *Journal of Applied Social Psychology*, vol. 20(4), pp. 276–302.
- Iyer, L., Mani, A., Mishra, P. and Topalova, P. (2012). 'The power of political voice: women's political representation and crime in India', *American Economic Journal: Applied Economics*, vol. 4(4), pp. 165–93.
- Johnson, B. (2014). 'Judges on trial: a reexamination of judicial race and gender effects across modes of conviction', *Criminal Justice Policy Review*, vol. 25(2), pp. 159–84.
- LaFree, G.D., Reskin, B.F. and Visser, C. (1985). 'Jurors' responses to victims' behavior and legal issues in sexual assault trials', *Social Problems*, vol. 32(4), pp. 389–407.
- Langbein, J. (1987). 'The English criminal trial jury on the eve of the French revolution', in (P. Schiappa, ed.), *The Trial Jury in England, France, Germany: 1700–1900*, pp. 13–39, Berlin: Duncker & Humblot.
- Lehmann, J. and Blair Smith, J. (2013). 'A multidimensional examination of jury composition, trial outcomes, and attorney preferences', available at: [http://www.uh.edu/~jlehman2/papers/lehmann\\_smith\\_jurycomposition.pdf](http://www.uh.edu/~jlehman2/papers/lehmann_smith_jurycomposition.pdf) (last accessed: 9 March 2018).
- McCammon, H.J. (2012). *The U.S. Women's Jury Movements and Strategic Adaptation: A More Just Verdict*, New York: Cambridge University Press.
- Mills, C. and Bohannon, W. (1980). 'Juror characteristics: to what extent are they related to jury verdicts?', *Judicature*, vol. 64(1), pp. 22–31.
- Moran, G. and Comfort, J. (1982). 'Scientific jury selection: sex as a moderator of demographic and personality predictors of impanelled felony juror behavior', *Journal of Personality and Social Psychology*, vol. 43(5), pp. 1052–63.
- Pearson, K. and Dancy, L. (2011). 'Elevating women's voices in congress: speech participation in the House of Representatives', *Political Research Quarterly*, vol. 64(4), pp. 910–23.
- Simon, R. (1967). *The Jury and The Defense of Insanity*, Boston, MA: Little, Brown and Company.
- Steffensmeier, D. and Herbert, C. (1999). 'Women and men policymakers: does the judge's gender affect the sentencing of criminal defendants?', *Social Forces*, vol. 77(3), pp. 1163–96.
- Thomas, C. (2010). 'Are juries fair?', Ministry of Justice Research Series No. 1/10.
- Thomas, C. and Balmer, N. (2007). 'Diversity and fairness in the jury system', Ministry of Justice Research Series No. 2/07.
- Thompson, W.C., Cowan, C.L., Ellsworth, P.C. and Harrington, J. (1984). 'Death penalty attitudes and conviction proneness: the translation of attitudes into verdicts', *Law and Human Behavior*, vol. 8(1), pp. 95–113.
- Volden, C., Wiseman, A.E. and Wittmer, D.E. (2013). 'When are women more effective lawmakers than men?', *American Journal of Political Science*, vol. 57(2), pp. 326–41.
- Wittmer, D.E. and Bouche, V. (2013). 'The limits of gendered leadership: policy implications of female leadership on women's issues', *Politics & Gender*, vol. 9(3), pp. 245–75.