Search for charged-lepton-flavour violation in Z-boson decays with the ATLAS detector

ATLAS Collaboration*

Leptons with essentially the same properties apart from their mass are grouped into three families (or flavours). The number of leptons of each flavour is conserved in interactions, but this is not imposed by fundamental principles. Since the formulation of the standard model of particle physics, the observation of flavour oscillations among neutrinos has shown that lepton flavour is not conserved in neutrino weak interactions. So far, there has been no experimental evidence that this also occurs in interactions between charged leptons. Such an observation would be a sign of undiscovered particles or a yet unknown type of interaction. Here the ATLAS experiment at the Large Hadron Collider at CERN reports a constraint on lepton-flavour-violating effects in weak interactions, searching for Z-boson decays into a τ lepton and another lepton of different flavour with opposite electric charge. The branching fractions for these decays are measured to be less than \(8.1 \times 10^{-6}\) (eτ) and \(9.5 \times 10^{-6}\) (μτ) at the 95% confidence level using 139 fb\(^{-1}\) of proton-proton collision data at a centre-of-mass energy of \(\sqrt{s} = 13\) TeV and 20.3 fb\(^{-1}\) at \(\sqrt{s} = 8\) TeV. These results supersede the limits from the Large Electron–Positron Collider experiments conducted more than two decades ago.

In the standard model of particle physics (SM), three lepton families (flavours) exist. The number of leptons of each family is conserved in weak interactions, and violation of this assumption is known as lepton flavour violation (LFV). No fundamental principles forbid LFV processes in the SM. The phenomenon of neutrino oscillations, where neutrinos (the neutral leptons) of one flavour transform into those of another\(^{1,4}\), indicates that neutrinos have mass and LFV processes do occur in nature. The mechanisms responsible for neutrinos acquiring mass and weak interactions violating lepton flavour conservation remain unknown. More experimental data are needed to constrain and guide possible generalizations of the SM explaining these phenomena.

An observation of LFV in charged-lepton interactions would be an unambiguous sign of new physics. In particular, decays of the Z boson into a light lepton (electron or muon) and a τ lepton at colliders are of experimental interest. The abundance of Z bosons produced at the Large Hadron Collider (LHC) offers the opportunity to strongly constrain potential LFV \(Z \rightarrow e\tau\) or \(Z \rightarrow \mu\tau\) interactions, in particular those proportional to the centre-of-mass energy of the decay\(^2\). Moreover, \(Z \rightarrow e\tau, \mu\tau\) decays are less constrained by low-energy experiments than \(Z \rightarrow e\mu\) decays. According to current knowledge, these decays can occur via neutrino mixing but are too rare to be detected. Only 1 in approximately \(10^{15}\) Z bosons would decay into a muon and a τ lepton\(^5\). An observation of such decays would therefore require new theoretical explanations. For example, theories predicting the existence of heavy neutrinos\(^6\) provide a fundamental understanding of the observed tiny masses and large mixing of SM neutrinos. In such theories, up to 1 in \(10^{8}\) Z bosons would be expected to undergo an LFV decay involving τ leptons. The ATLAS experiment can test the predictions of such theories by observing or setting even more stringent constraints on LFV Z-boson decays.

Constraints on the branching fractions (\(B\)) of the LFV decays of the Z boson involving a τ lepton have been set by the experiments at the Large Electron–Positron Collider (LEP):

- \(B(Z \rightarrow e\tau) < 9.8 \times 10^{-6}\) (ref. \(^{10}\))
- \(B(Z \rightarrow \mu\tau) < 1.2 \times 10^{-5}\) (ref. \(^{11}\)) at the 95% confidence level (CL). The ATLAS experiment\(^2\) at the LHC has set constraints \(B(Z \rightarrow e\tau) < 5.8 \times 10^{-5}\) at 95% CL, using part of the Run 2 data and \(B(Z \rightarrow \mu\tau) < 1.3 \times 10^{-5}\) using the Run 1 data and a subset of the Run 2 data\(^{11}\).

This work uses proton–proton (pp) collision data collected by the ATLAS experiment during Run 2 of the LHC, containing about eight billion Z-boson decays. Only events with a τ lepton that decays hadronically are considered. Neural network (NN) classifiers are used in a novel way for optimal discrimination of signal from background, and to achieve improved sensitivity in the search for LFV effects in the data using a binned maximum-likelihood fit. The result for the \(\mu\tau\) channel is combined with a previous LHC Run 1 result to further improve the sensitivity. These results set constraints on LFV Z-boson decays involving τ leptons that supersede the most stringent ones set by the LEP experiments more than two decades ago.

The ATLAS experiment and data sample

To record and analyse the LHC pp collisions, the ATLAS experiment uses a multipurpose particle detector with a forward–backward symmetric cylindrical geometry and a near 4π coverage in solid angle\(^{12,14,15}\). It consists of an inner tracking detector surrounded by a superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon spectrometer.

The search uses the complete dataset of pp collision events at a centre-of-mass energy of \(\sqrt{s} = 13\) TeV collected by the ATLAS experiment during LHC Run 2. This dataset was recorded using single-electron or single-muon triggers\(^{16}\) and corresponds to an integrated luminosity of 139 fb\(^{-1}\). For the search in the \(\mu\tau\) channel, the results are combined with those of a previous similar search using pp collisions at \(\sqrt{s} = 8\) TeV during LHC Run 1, corresponding to an integrated luminosity of 20.3 fb\(^{-1}\) (ref. \(^{17}\)).

Candidates for electrons\(^{18}\), muons\(^{19}\), jets\(^{20–22}\), and visible decay products of hadronic τ-lepton decays (\(τ_{had-nu}\))\(^{23,24}\) are reconstructed from energy deposits in the calorimeters and charged-particle tracks measured in the inner detector and the muon spectrometer.

*A list of authors and their affiliations appears online.
Electron candidates are required to pass the Medium likelihood-based identification requirement and have pseudorapidity $|\eta| < 1.37$ or $1.52 < |\eta| < 2.47$. Muon candidates are required to pass the Medium identification requirement and have $|\eta| < 2.5$. Both the electron and muon candidates must have transverse momentum $p_T > 30 \text{ GeV}$ and satisfy the Tight isolation requirement. The lower bounds on the electron and muon transverse momenta are driven by the acceptance of the trigger selection.

Quark- or gluon-initiated particle showers (jets) are reconstructed using the anti-$k_T$ algorithm with the radius parameter $R = 0.4$. Jets fulfilling $p_T > 20 \text{ GeV}$ and $|\eta| < 2.5$ are identified as containing $b$ hadrons if tagged by a dedicated multivariate algorithm.

The $\tau_{\text{had-vis}}$ candidates are reconstructed from jets with $p_T > 10 \text{ GeV}$, $|\eta| < 1.37$ or $1.52 < |\eta| < 2.5$, and one or three associated tracks, referred to as ‘1-prong’ (1P) and ‘3-prong’ (3P), respectively. The $\tau_{\text{had-vis}}$ identification is performed by a recurrent NN algorithm, which uses calorimetric shower shapes and tracking information to discriminate true $\tau_{\text{had-vis}}$ candidates from fake candidates from quark- or gluon-initiated jets. The $\tau_{\text{had-vis}}$ candidates are required to pass the Tight identification selection, which has an efficiency of 60% (45%) for true 1P (3P) $\tau_{\text{had-vis}}$ candidates, constant in the $\tau_{\text{had-vis}}$ candidates' transverse momentum, and a misidentification rate of 1 in 70 (700) for fake 1P (3P) candidates in dijet events. Dedicated multivariate algorithms are used to further discriminate between $\tau_{\text{had-vis}}$ and electrons, and to calibrate the $\tau_{\text{had-vis}}$ energy. The $\tau_{\text{had-vis}}$ candidate with the largest $p_T$ in each event is the selected candidate and is required to have $p_T > 25 \text{ GeV}$. Based on simulation, in $Z \rightarrow \ell\ell$ decays, the $\tau_{\text{had-vis}}$ candidate is expected to be correctly selected 98% of the time.

The missing transverse momentum ($E_T^{\text{miss}}$) is calculated as the negative vectorial sum of the $p_T$ of all fully reconstructed and calibrated physics objects. The calculation also includes inner detector tracks that originate from the vertex associated with the hard-scattering process but are not associated with any of the reconstructed objects. The missing transverse momentum is the best proxy for the total transverse momentum of undetected particles (in particular neutrinos) in an event.

**Search strategy**

The $Z \rightarrow \ell\ell \rightarrow \tau\tau$ signal events have a number of key features that can be exploited to separate them from the SM background events. The signal events are characterized by their unique final state, which has exactly one $\ell$ and one $\tau$ lepton, with the invariant mass of the pair being compatible with the $Z$-boson mass. The $\ell$ and $\tau$ leptons carry opposite electric charges, which use calorimetric shower shapes and tracking information to discriminate true $\tau_{\text{had-vis}}$ candidates from fake candidates from quark- or gluon-initiated jets. The $\tau_{\text{had-vis}}$ candidate is required to pass the Tight identification selection, which has an efficiency of 60% (45%) for true 1P (3P) $\tau_{\text{had-vis}}$ candidates, constant in the $\tau_{\text{had-vis}}$ candidates' transverse momentum, and a misidentification rate of 1 in 70 (700) for fake 1P (3P) candidates in dijet events. Dedicated multivariate algorithms are used to further discriminate between $\tau_{\text{had-vis}}$ and electrons, and to calibrate the $\tau_{\text{had-vis}}$ energy. The $\tau_{\text{had-vis}}$ candidate with the largest $p_T$ in each event is the selected candidate and is required to have $p_T > 25 \text{ GeV}$. Based on simulation, in $Z \rightarrow \ell\ell$ decays, the $\tau_{\text{had-vis}}$ candidate is expected to be correctly selected 98% of the time.

The missing transverse momentum ($E_T^{\text{miss}}$) is calculated as the negative vectorial sum of the $p_T$ of all fully reconstructed and calibrated physics objects. The calculation also includes inner detector tracks that originate from the vertex associated with the hard-scattering process but are not associated with any of the reconstructed objects. The missing transverse momentum is the best proxy for the total transverse momentum of undetected particles (in particular neutrinos) in an event.

The major background contributions for this search are as follows: lepton-flavour-conserving $Z \rightarrow \ell\ell \rightarrow \ell\ell$ decays, where one of the $\ell$ leptons decays leptonically and the other hadronically; $Z \rightarrow \ell\ell$ decays, where one of the light leptons is misidentified as the $\tau_{\text{had-vis}}$ candidate; events with a quark- or gluon-initiated jet that is misidentified as the $\tau_{\text{had-vis}}$ candidate. The last of these are hereafter referred to as events with ‘fakes’ and are mostly $W(\rightarrow \ell\nu)$ + jets events and purely hadronic multijet events. Other SM processes with a real $\tau_{\text{had-vis}}$ final state, such as decays of a top–antitop-quark pair, two gauge bosons or a Higgs boson, and those with a real $\tau_{\text{had-vis}}$ and a jet misidentified as a light lepton, such as $W(\rightarrow \tau\nu)$ + jets, are considered, although their contribution to the overall background is minor.

The signal and background events are separated by using a set of event selection criteria that help to define a signal-enhanced sample, referred to as the signal region (SR). The main selection criteria are listed in Table 1 and will be explained in the following. They are primarily based on the multiplicity of reconstructed particle candidates and the event topology, in particular the transverse masses ($m_T$), which are defined as

$$m_T(X, E_T^{\text{miss}}) = \sqrt{2p_T(X)E_T^{\text{miss}}(1 - \cos(\phi_X - \phi_{E_T^{\text{miss}}}))}$$

where $X$ is either a light lepton or a $\tau_{\text{had-vis}}$ candidate and $\phi$ denotes the azimuthal angle. A schematic of the expected signal and background topologies is described in Extended Data Figs. 1 and 2.

<table>
<thead>
<tr>
<th>Main selection criteria</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least one $\tau_{\text{had-vis}}$ candidate</td>
<td>Select events with $\tau$–$\tau$ pair</td>
</tr>
<tr>
<td>Exactly one isolated light lepton</td>
<td>candidate</td>
</tr>
<tr>
<td>Opposite-sign charged $\ell$–$\tau_{\text{had-vis}}$ pair</td>
<td></td>
</tr>
<tr>
<td>$m_T(\tau_{\text{had-vis}}, E_T^{\text{miss}}) &lt; 35 \text{ GeV}$</td>
<td>Reject $Z \rightarrow \tau\tau$ and $W +$ jets events</td>
</tr>
<tr>
<td>$m_{\ell\ell}(\ell_1, \ell_2; \tau_{\text{had-vis}}) &gt; 60 \text{ GeV}$</td>
<td>Invariant mass of the $\ell$–$\tau_{\text{had-vis}}$ pair. Reject events incompatible with $\ell$–$\tau$ pairs from $Z$-boson decays</td>
</tr>
<tr>
<td>No tagged $b$-hadron jets</td>
<td>Reject if and single-top-quark events</td>
</tr>
<tr>
<td>Combined NN output $&gt; 0.1$ for events with 1P (3P) $\tau_{\text{had-vis}}$ candidates</td>
<td>Reject background-like events</td>
</tr>
<tr>
<td>NN (optimized for signal versus $Z \rightarrow \ell\ell$ output $&gt; 0.2$)</td>
<td>Ensure orthogonal region for correcting $Z \rightarrow \ell\ell$ simulation ($\ell$ misidentified as 1P $\tau_{\text{had-vis}}$ candidate, see section ‘Signal and background predictions’)</td>
</tr>
</tbody>
</table>

The outputs from the individual NNs are numbers between 0 and 1 that reflect the probability for an event to be a signal event; they are combined into a final discriminant, hereafter referred to as the ‘combined NN output’. The combination is parameterized by weights associated with each individual NN and optimized for discrimination among various background processes distributed differently along the range of combined NN output values, as detailed in Methods. This allows the maximum-likelihood fit to determine the background contributions more precisely, which ultimately improves the sensitivity.

Events classified by the NNs as being background-like are excluded from the SR, as indicated in Table 1. The signal acceptance times selection efficiency in the SR is 2.7% for the $\tau\tau$ channel and 3.0% for the $\mu\tau$ channel, as determined from simulated signal samples.
Signal and background predictions

Predictions for signal and background contributions to the event yield and kinematic distributions in the SR are based partly on Monte Carlo (MC) simulations and partly on the use of data in regions that are enriched in background events and do not overlap with the SR.

The signal events were simulated using PYTHIA $8^{28}$ with matrix elements calculated at leading order (LO) in the strong coupling constant ($\alpha_s$). Parameter values for initial-state radiation, multiparton interactions and beam remnants were set according to the A14 set of tuned parameters (tune)$^{29}$ with the NNPDF 2.3 LO parton distribution function (PDF) set$^{30}$. Nominal signal samples were generated with a parity-conserving $Z\ell\tau$ vertex and unpolarized $\tau$ leptons. Scenarios where the decays are maximally parity-violating were considered by reweighting the simulated events using TAUSPINNER$^{31}$. The event weight was computed as the probability of occurrence of each generated signal event, based on its kinematics, when assuming a specific $\tau$-polarization state (left-handed or right-handed).

Background $Z \rightarrow \tau\tau$ events were simulated with the SHERPA 2.2.1$^{32}$ generator using the NNPDF 3.0 NNLO PDF set$^{33}$ and next-to-leading-order (NLO) matrix elements for up to two partons.
Table 2 | Summary of the uncertainties and their impacts on the measured branching fraction $B(Z \rightarrow \ell\tau)$

<table>
<thead>
<tr>
<th>Source of uncertainty</th>
<th>Uncertainty on $\frac{B(Z \rightarrow \ell\tau)}{B(Z \rightarrow \ell\tau)}$ (x10^{-4})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>±3.5</td>
</tr>
<tr>
<td>Systematic</td>
<td>±2.3</td>
</tr>
<tr>
<td>$r$ leptons</td>
<td>±1.9</td>
</tr>
<tr>
<td>Energy calibration</td>
<td>±1.3</td>
</tr>
<tr>
<td>Jet rejection</td>
<td>±0.3</td>
</tr>
<tr>
<td>Electron rejection</td>
<td>±1.3</td>
</tr>
<tr>
<td>Light leptons</td>
<td>±0.4</td>
</tr>
<tr>
<td>$E_{\text{miss}}^\tau$, jets and flavour tagging</td>
<td>±0.6</td>
</tr>
<tr>
<td>Z-boson modelling</td>
<td>±0.7</td>
</tr>
<tr>
<td>Luminosity and other minor backgrounds</td>
<td>±0.8</td>
</tr>
<tr>
<td>Total</td>
<td>±4.1</td>
</tr>
</tbody>
</table>

The statistical uncertainties include those in the determination of the yields of the events with fakes and from $Z \rightarrow \ell\tau$ or $Z \rightarrow \ell\ell$ decays. The uncertainties related to light leptons include those in the trigger, reconstruction, identification and isolation efficiencies, as well as energy calibrations. The uncertainties related to jets and $E_{\text{miss}}^\tau$ include those in energy calibration and resolution. The uncertainties related to the Z-boson modelling include those in the correction of the simulated transverse momentum and the measured production cross-section of the Z boson.

and LO matrix elements for up to four partons, calculated with the COMIX4 and OPENLOOPS5–7 libraries. They were matched with the SHERPA parton shower8 using the MEPS@NLO prescription9–11, with the default SHERPA tune. This tune follows the recommendations of the SHERPA authors. Background $Z \rightarrow \ell\ell$ events were simulated using the POWHEG-BOX12 generator with NLO matrix elements and interfaced to PYTHIA8 to model the parton showers, hadronization and underlying events. All MC samples include a detailed simulation of the ATLAS detector with parton showers, hadronization and underlying events. All MC samples include a detailed simulation of the ATLAS detector with parton showers, hadronization and underlying events. They were matched with the SHERPA parton shower8 using the MEPS@NLO prescription9–11, with the default SHERPA tune. This tune follows the recommendations of the SHERPA authors.

The simulation of Z-boson production is improved with a correction derived from measurements in data. The simulated $p_T$ spectra of the Z boson are reweighed to match the unfolded distribution measured by ATLAS in ref. 46. This improves the predictions of signal, $Z \rightarrow \tau\tau$ and $Z \rightarrow \ell\ell$ events, which are simulated at different $p_T$ ranges in the SR or CRZ$\tau\tau$. The modelling of the estimated background is validated using data in the CRZ$\tau\tau$ and in a region similar to the SR, but with events that have same-sign charged $\ell^±$, $\tau^±$, $E_{\text{miss}}^\tau$, pairs, as shown in Fig. 1.

### Constraints on $B(Z \rightarrow \ell\tau)$

A statistical analysis of the selected events is performed to assess the presence of LFV signal events. The statistical analysis method is detailed in Methods. A simultaneous binned maximum-likelihood fit to the combined NN output in the SR and $m_\ell(E_{\text{miss}}^\tau)$ in the CRZ$\tau\tau$ is used to constrain uncertainties in the models and extract evidence of a possible signal. The fit is performed independently for the $\ell$ and $\mu$ channels. Events with 1P and 3P $\tau_{\text{had-vis}}$ candidates are considered separately. Hypothesis tests, in which a log-likelihood ratio is used as the test statistic, are used to assess the compatibility between the background and signal models and the data.

There are four unconstrained parameters in the fits: two of them determine the overall yields of events with fake 1P $\tau_{\text{had-vis}}$ or 3P $\tau_{\text{had-vis}}$ candidates.
In the case of no significant deviations from the SM background, exclusion limits are set using the CLS method. For systematic uncertainties in the signal and background predictions, one determines $\sigma_\text{fit}$ times the overall acceptance and reconstruction efficiency of the $1\tau$ final state in $Z \to \tau\tau$ and signal events, and the last one, the parameter of interest, determines the LFV branching fraction $B(Z \to \ell\tau)$ by modifying an arbitrary pre-fit signal yield.

Constrained parameters are also introduced to account for systematic uncertainties in the signal and background predictions. In the case of no significant deviations from the SM background, exclusion limits are set using the CLS method. Systematic uncertainties in this search include uncertainties in simulated events in the modelling of trigger, reconstruction, identification and isolation efficiencies, as well as energy calibrations and resolutions of reconstructed objects. Conservative theory uncertainties ranging between 4% and 20% are also assigned to the predicted cross-sections used for the estimation of minor background processes. These uncertainties are not assigned to events with fakes or $Z$-boson decays, whose yields are determined from data. These events constitute only a small fraction of the background events in the SR. The dominant uncertainties in this search are those in the overall yields of events with fakes, which are predominantly of statistical nature, and those in the $\tau_{\text{had}}\mu$ energy calibration, which are independent between 1P and 3P $\tau_{\text{had}}\mu$ candidates.
and constrained by the fit of the collinear mass spectrum to the data in the \( \text{CR}_{Z\tau} \). A summary of the uncertainties and their impact on the best-fit LFV branching fraction is provided in Table 2, which shows that the sensitivity of the search is primarily limited by the available amount of data.

The best-fit observed and expected distributions of the combined NN output in the SR are shown in Fig. 2. The best-fit yields of \( Z \to \tau\tau \) and events with fakes are close to the pre-fit predicted values and are determined with a relative precision of 2–4%. Table 3 shows the best-fit expected background and signal yields and the observed number of events in the SR of the \( \tau \) and muon channels with an additional requirement of a combined NN output greater than 0.7 to consider the most signal-like events.

The best-fit amount of \( Z \to \ell\tau \) signal corresponds to the branching fractions \( B(Z \to \tau\tau) = (0.1 \pm 3.5(\text{stat}) \pm 2.3(\text{syst})) \times 10^{-6} \) and \( B(Z \to \mu\tau) = (4.3 \pm 2.8(\text{stat}) \pm 1.6(\text{syst})) \times 10^{-6} \). The positive best-fit value of \( B(Z \to \mu\tau) \) is related to a small excess of observed events relative to the background-only hypothesis. This excess has a significance of 0.9 standard deviations when the events with 1P and 3P \( \tau_{\text{had}} \) candidates are fitted simultaneously.

No statistically significant deviation from the SM prediction is observed, and upper limits on the LFV branching fractions are set. For the \( \mu\tau \) channel, a more stringent upper limit is set by combining the likelihood function of the presented measurement and a similar measurement done with ATLAS Run 1 data\(^7\). Systematic uncertainties from the two measurements are considered uncorrelated in the combined likelihood function. The upper limits are shown in Table 4 for LFV decays with different assumptions about the \( r \) polarization state.

In conclusion, these results from the ATLAS experiment at the LHC set stringent constraints on LFV \( Z \)-boson decays involving \( \tau \) leptons (using only their hadronic decays), superseding the most stringent ones set by the LEP experiments more than two decades ago. The precision of these results is mainly limited by statistical uncertainties.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41567-021-01225-z.

Received: 7 October 2020; Accepted: 11 March 2021; Published online: 1 July 2021

References

Methods

Neural network classifiers. Several binary NN classifiers are trained for both the $e\tau$ and $\mu\tau$ channels to discriminate signal from the three major backgrounds: $W+\text{jets}, Z\rightarrow\tau\tau$ and $Z\rightarrow\ell\ell$. They are referred to as $\text{NN}_{bwb}$, $\text{NN}_{wb\tau}$ and $\text{NN}_{w\tau\tau}$, respectively.

The NNs are trained using simulated events selected with the same criteria as those used in the SR, except that the cuts on $m_\ell(E_t)$ and the event output are omitted, and real $\tau_{\text{had}}$ candidates from $Z\rightarrow\ell\ell$ and $Z\rightarrow\tau\tau$ are required to pass less stringent identification criteria so as to increase the training sample size. For the $Z\rightarrow\ell\ell$ process, only events where the $\tau_{\text{had}}$ candidate is a misidentified light lepton are used. For the $W+\text{jets}$ process, jets misidentified as $\tau$ are modelled by simulations. Different NNs are trained separately for $e\tau$ and $\mu\tau$ events as well as for events with $1\text{P}$ or $3\text{P}\tau_{\text{had}}$ candidates. To increase the signal sample size, the $Z\rightarrow\ell\ell$ and $Z\rightarrow\mu\mu$ samples are combined and used for training in both channels, assuming equivalent event topology when exchanging $e$ and $\mu$. Owing to the low expected yield of $Z\rightarrow\ell\ell$ events with $3\text{P}\tau_{\text{had}}$ candidates, no classifier is trained to discriminate them from background.

A mixture of low-level and high-level kinematic variables are used as input to the NNs. The low-level variables include the four-momenta of the reconstructed $\ell$ (refs. 51, 52, 53), $\tau_{\text{had}}$ candidate, and $E_T^{\text{miss}}$ (refs. 14, 54). To remove known spatial symmetries for optimal training, the low-level variables are transformed in a way that preserves the Lorentz invariance before they are fed into the NNs. The transformation consists of the following steps: first, the $\ell^{-}\tau_{\text{had}}^{-}E_T^{\text{miss}}$ system is boosted in a direction in the plane transverse to the beam line such that the total transverse momentum of the system is zero; the system is then rotated about the $Z$ axis by 180°. After the transformation, only six independent non-vanishing components are left (the $\tau_{\text{had}}$ candidate is assumed to have zero rest mass), which are the inputs to the NNs.

The high-level variables include $\Delta\alpha$, which is a kinematic discriminant defined as

$$\Delta\alpha = \frac{m_T^2 - m_\ell^2}{2p(E_T^{\text{miss}}) \times p(\tau_{\text{had}}-\ell)} - \frac{p(E_T^{\text{miss}})}{p(\tau_{\text{had}}-\ell)}$$

where $m_\ell$ and $m_T$ are the nominal masses of the $Z$ boson and $\ell$ lepton, respectively, and $\rho$ denotes four-momentum. It is specifically defined to test the assumptions that the missing momentum of the event is collinear with the $\tau_{\text{had}}$ candidate, and that the $Z$ and light leptons in the event are decay products of an on-shell $Z$ boson. For a signal event, where these assumptions are approximately true, it is expected that $\Delta\alpha = 0$. Meanwhile, for an SM background event, the value is expected to deviate from zero in general. The other high-level variables are the invariant mass of the $\ell^{-}\tau_{\text{had}}^{-}E_T^{\text{miss}}$ system, the collinear mass $m_\ell(E, \tau)$ and the invariant mass of the light lepton and the trick associated with the $\tau_{\text{had}}$ candidate (only used by the $Z\rightarrow\ell\ell$ classifier).

The training and optimization of the NN classifiers are performed using the open-source software package KERAS. All of the NNs used in the analysis share the same architecture. Each NN consists of an input layer, two hidden layers of 20 nodes each, and an output layer with a single node. Each layer is fully connected to the neighbouring layers. Low-level and high-level variables are treated in the same way in the input layer. The hidden-layer nodes use rectified linear activation functions, while the output node uses a sigmoid activation function. The NNs are trained using the Adam algorithm to optimize the binary cross entropy. All the NNs are trained with a batch size of 256 and 200 epochs. The number of hidden layers, the number of nodes per layer, the training batch size and the learning rate parameter of the optimizer are simultaneously chosen by maximizing the area under the expected receiver operating characteristic curve. The optimization is done with a grid scan. No regularization or dropout is added, and no sign of overtraining is observed. For other configurations and hyperparameters that have not been mentioned, the default settings in KERAS 1.1.0 are used.

Each NN classifier outputs a score between 0 and 1 for each event, where a higher score indicates that the event is more signal-like. The output scores from the different classifiers are combined into the final discriminant (combined NN output) using the formula

$$\text{Combined NN output} = 1 - \sqrt{\frac{\sum w_i \times (1 - \text{NN}_i \text{output})^2}{\sum w_i}}$$

where $b = \text{Wjets}, Z\tau, Z\ell\ell$ and $w_i$ are constant parameters. Output scores for events with $1\text{P}\tau_{\text{had}}$ candidates and those with $3\text{P}\tau_{\text{had}}$ candidates are combined separately. The summarization is over $\text{Wjets}, Z\tau$ and $Z\ell\ell$ for events with $1\text{P}\tau_{\text{had}}$ candidates, and only over $\text{Wjets}$ and $Z\tau$ for events with $3\text{P}\tau_{\text{had}}$ candidates. By construction, the combined NN output ranges between 0 and 1, where 0 represents the most background-like (and 1 the most signal-like) event possible.

The choice of values of $w_i$ affects the expected sensitivity of the analysis because they change how events from the different background processes are distributed along the range of combined NN output values, and thus impact the ability of the binned maximum-likelihood fit to determine the background contributions. The values of $w_i$ are chosen with a grid scan to minimize the expected upper limit on the branching fraction in the absence of a signal. The chosen values have the form $w_i = \text{NN}_i\tau_{\text{had}}(2\text{P}) = 1.0 \pm 0.033$. As could be expected, the optimized weights loosely reflect the impact of the uncertainties in the corresponding backgrounds on the determination of the signal branching fraction.

Maximum-likelihood fit. Binned maximum-likelihood fits are implemented using the statistical analysis packages ROOFIT, BOOSTSTAT and HISTFITTER.

The expected binned distributions of the combined NN output in the SR and the collinear mass in the CRZ$\tau\tau$ are fit to data to extract evidence of signal events. Fitting the data in the CRZ$\tau\tau$ and in part of the SR with low combined NN output values (where no signal is expected) benefits the overall sensitivity to the signal, because it reduces the uncertainties of the background model in the CRZ$\tau\tau$ region, where most of the signal is expected. Owing to the differences in background composition, acceptance and efficiencies, regions with $1\text{P}$ and $3\text{P}\tau_{\text{had}}$ candidates are fit separately but simultaneously. The probabilities of compatibility between the data and the background-only or background-plus-signal hypotheses are assessed using the modified frequentist CL$_S$ method and exclusion upper limits on $B(Z\rightarrow\ell\tau)$ are set by the inversion of these hypothesis tests.

The background-plus-signal model has four unconstrained parameters before the fit. Two of the parameters determine the overall yields of events with $1\text{P}$ and $3\text{P}$ fakes separately. A third parameter determines $\sigma_\tau$, the overall acceptance and reconstruction efficiency of events with a $1\text{P}\tau_{\text{had}}$ final state. It is applied to the normalizations of both the signal and $Z\rightarrow\ell\ell$ events to ensure that the same $\sigma_\tau$ normalization is estimated for both processes. The last unconstrained parameter is the parameter of interest $\mu$ which controls the normalization of signal events. Given the similarity between the signal and $Z\rightarrow\ell\ell$ for events with $1\text{P}$ and $3\text{P}\tau_{\text{had}}$ final states and that both processes are estimated with the same $\sigma_\tau$, acceptance and efficiency corrections, this choice of parameterization reduces the impact on the determined $B(Z\rightarrow\ell\tau)$ from detector effects and uncertainties in predicting $\sigma_\tau$. The parameter of interest represents $B(Z\rightarrow\ell\tau)$ on the branching fraction in the absence of a signal. The chosen values have the form $\text{NN}_i\tau_{\text{had}}(2\text{P}) = 1.0 \pm 0.033$. As could be expected, the optimized weights loosely reflect the impact of the uncertainties in the corresponding backgrounds on the determination of the signal branching fraction.

Data availability

The experimental data that support the findings of this study are available in HEPData with the identifier https://www.hepdata.net/record/963900. The ATLAS software is available at the following link: https://gitlab.cern.ch/atlas/athena.

References

49. Chollet, F. et al. Keras (Keras, 2015); https://keras.io
54. ATLAS Collaboration ATLAS Computing Acknowledgments ATL-SOFT-PUB-2020-001 (CERN, 2020); https://cds.cern.ch/record/2717821
Acknowledgements
We acknowledge our late colleague, Olga Igonkina (1973–2019), for inspiring and driving this and other searches for lepton flavour violation within the ATLAS experiment. Her curiosity and intelligence remain an inspiration to the ATLAS Collaboration. We thank CERN for the very successful operation of the LHC, as well as the support staff from our institutions, without whom ATLAS could not be operated efficiently. We acknowledge the support of ANPCyT, Argentina; YerPhI, Armenia; ARC, Australia; BMWF and FWF, Austria; ANAS, Azerbaijan; STFC, Belarus; CNPq and FAPESP, Brazil; NSERC, NRC and CFI, Canada; CERN; ANID, Chile; CAS, MOST and NSFC, China; GOLCIENTRIAS, Colombia; MSMT CR, MPO CR and VSC CR, Czech Republic; DURDF and DURSNC, Denmark; IN2P3-CNRS and CEA-DRF/Irfu, France; SRNSFG, Georgia; BMBF, HGF and MPG, Germany; GRS, Greece; RGC and Hong Kong SAR, China; ISF and Benoziyo Center, Israel; INFN, Italy; MEXT and JSPS, Japan; CNRST, Morocco; NWO, Netherlands; RCN, Norway; MNiSW and NCN, Poland; FCT, Portugal; MNE/IFA, Romania; MES of Russia and NRC KI, Russia Federation; JINR; MESTD, Serbia; MSMR, Slovakia; ARRS and MIZŠ, Slovenia; DST/NRF, South Africa; MICINN, Spain; SRC and Wallenberg Foundation, Sweden; SERI, SNSF and Cantons of Bern and Geneva, Switzerland; MOST, Taiwan; TAEK, Turkey; STFC, UK; DOE and NSF, United States. In addition, individual groups and members have received support from BCKDF, CANARIE, Compute Canada and CRC, Canada; ERC, ERDF, Horizon 2020, Marie Skłodowska-Curie Actions and COST, EU; Investissements d'Avenir Labex, Investissements d'Avenir Idex and ANR, France; DFG and AvH Foundation, Germany; Herakleitos, Thales and Aristeia programmes co-financed by EU-ESF and the Greek NSRF; Greece; BSF-NSF and GIF, Israel; La Caixa Banking Foundation, CERCA Programme Generalitat de Catalunya and PROMETEO and Geit Programmes Generalitat Valenciana, Spain; Goran Gustafssons Stiftelse, Sweden; The Royal Society and Leverhulme Trust, UK. The crucial computing support from all WLCG partners is acknowledged gratefully, in particular from CERN, the ATLAS Tier-1 facilities at TRIUMF (Canada), NDGF (Denmark, Norway, Sweden), CC-IN2P3 (France), KIT/GridKA (Germany), INFN-CNAF (Italy), NL-T1 (Netherlands), PIC (Spain), ASGC (Taiwan), RAL (UK) and BNL (USA), the Tier-2 facilities worldwide and large non-WLCG resource providers. Major contributors of computing resources are listed in ref. 44.

Author contributions
All authors contributed to the publication, being variously involved in the design and the construction of the detectors, in writing software, calibrating subsystems, operating the detectors and acquiring data, and finally analysing the processed data. The ATLAS Collaboration members discussed and approved the scientific results. The manuscript was prepared by a subgroup of authors appointed by the collaboration and subject to an internal collaboration-wide review process. All authors reviewed and approved the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41567-021-01225-z.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41567-021-01225-z.
Correspondence and requests for materials should be addressed to G.A.

Peer review information Nature Physics thanks Michael Schmidt, Roger Wolf and Scott Yost for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
ATLAS Collaboration


Extended Data Fig. 1 | Schematic representation of a typical event selected in the SR. The topology as seen in the plane transverse to the beam line is shown. (a) A signal $Z \rightarrow \ell \tau$ event. (b) A $Z \rightarrow \tau \tau$ event. (c) A $W$+jets event. The green arrows represent reconstructed light leptons ($\ell$). The blue triangles represent the $\tau_{\text{had-vis}}$ candidates. The light blue dashed lines represent neutrinos that escape detection and are reconstructed as (part of) the missing transverse momentum of the event.
Extended Data Fig. 2 | Distributions of $m_T(\tau_{\text{had}}-\tau_{\text{vis}}, E_T^{\text{miss}})$ versus $m_T(\mu, E_T^{\text{miss}})$ of events selected in the SR. (a) Simulated $Z \rightarrow \mu\tau$ events. (b) Simulated $Z \rightarrow \tau\tau$ events. (c) Events measured in data in regions where quark- or gluon-initiated jets are misidentified as $\tau_{\text{had}}$ candidates (events with jet $\rightarrow \tau_{\text{had}}$ fakes, see ‘Signal and background predictions’ section) in the $\mu\tau$ final state. The colour map represents the fraction of events in each bin.