Gravitational-wave detections provide a novel way to determine the Hubble constant, which is the current rate of expansion of the Universe. This ‘standard siren’ method, with the absolute distance calibration provided by the general theory of relativity, was used to measure the Hubble constant using the gravitational-wave detection of the binary neutron-star merger, GW170817, by the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo, combined with optical identification of the host galaxy NGC 4993. This independent measurement is of particular interest given the discrepancy between the value of the Hubble constant determined using type Ia supernovae via the local distance ladder and the value determined from cosmic microwave background observations; these values differ by about 3σ. Local distance ladder observations may achieve a precision of one per cent within five years, but at present there are no indications that further observations will substantially reduce the existing discrepancies. Here we show that additional gravitational-wave detections by LIGO and Virgo can be expected to constrain the Hubble constant to a precision of approximately two per cent within five years and approximately one per cent within a decade. This is because observing gravitational waves from the merger of two neutron stars, together with the identification of a host galaxy, enables a direct measurement of the Hubble constant independent of the systematics associated with other available methods. In addition to clarifying the discrepancy between existing low-redshift (local ladder) and high-redshift (cosmic microwave background) measurements, a precision measurement of the Hubble constant is of crucial value in elucidating the nature of dark energy.

We explore the expected constraints on the Hubble constant (H₀) from gravitational-wave standard sirens. The gravitational-wave data provide a direct measurement of the luminosity distance to the source, but the redshift must be determined independently. We consider gravitational-wave events both with (‘counterpart’) and without (‘statistical’) direct electromagnetic measurements of the source redshift, and carry out an end-to-end simulation of the H₀ measurement from a simulated dataset consisting of 30,000 binary neutron star (BNS) mergers and 60,000 binary black hole (BBH) mergers. We include realistic measurement uncertainties, galaxy peculiar velocities and selection effects in our analysis.

We anticipate that most, if not all, BNS mergers detected in gravitational waves will have an electromagnetic counterpart (for example, from associated isotropic kilonova emission) that will allow a unique host galaxy identification. Assuming that the BNS population is similar to the population of short γ-ray bursts, we expect the typical offset between a kilonova and its associated host galaxy to be no more than 100 kpc. Since Advanced LIGO–Virgo BNS detections will be within 400 Mpc, it will be possible to identify host galaxies as faint as 0.003L* (apparent magnitudes <23), where L* is the B-band Schechter function parameter (see Methods), with modest observational resources. We find that in this counterpart case, the fractional H₀ uncertainty will scale roughly as 15% / √N, where N is the number of BNS mergers detected by the LIGO-Hanford, LIGO-Livingston and Virgo network (HLV). Throughout, we quote fractional H₀ measurement uncertainties defined as half the width of the symmetric 68% credible interval divided by the median. If the Kamioka Gravitational Wave Detector (KAGRA) and LIGO-India join the detector network (HLVJ), this convergence improves slightly to 13% / √N, because a five-detector network tends to provide better measurements of the source inclination, and therefore distance, owing to the improved polarization information.

We note that the representative fractional H₀ measurement uncertainty (15% for the three-detector network, and 13% for the five-detector network), is smaller than the typical width of the H₀ measurement from an individual event (GW170817 provided an unusually tight measurement; see Extended Data Fig. 1). This is because for a single event, the H₀ posterior probability density function is a highly non-Gaussian function; the distance–inclination degeneracy leads to long tails up to large distances (and low H₀ values) for edge-on sources, and tails in the opposite direction for face-on sources. Combining these asymmetric distributions leads to a 1 / √N convergence rate. For example, we may get lucky in the first few events and get an unusually good H₀ measurement (GW170817 is an excellent example of this), after which we will converge more slowly than 1 / √N for some time as we detect average events. After about 20 events, however, we will have a sufficient statistical sample of detections to have converged to a representative H₀ for the population. At this point, the combined H₀ measurement approaches a Gaussian distribution and we reach the expected 1 / √N behaviour.

To predict how the H₀ measurement improves with time, we consider the BNS rates inferred from GW170817, Design HLV and Design HLVJI, respectively. The merger rate of BNS detections per year in O3 HLV, to 32 and 39 detections per year in Design HLV and Design HLVJI, respectively. The merger rate measurement as BNS detections with unique host galaxies (and associated redshifts) are accumulated. We start with a 15% prior measurement on H₀ representing the constraint from GW170817; we approximate this by a Gaussian distribution. We then show the improvement in H₀ measurement as BNS detections with unique host galaxies (and associated redshifts) are accumulated. We start with a 15% prior measurement on H₀ representing the constraint from GW170817; we approximate this by a Gaussian distribution.
We find that, if it is possible to independently measure a unique redshift for all BNS events, the fractional uncertainty on $H_0$ will reach 2% (at the 1σ level) by the end of two years of HLV at design sensitivity (in about 2023; corresponding to about 50 events), sufficient to arbitrate the current tension between local and high-$z$ measurements of $H_0$. After about 100 BNS events, gravitational-wave standard sirens would provide a 1% determination of $H_0$. This is expected to happen after about two years of operation of the full HLVJI network (around 2026), but given the rate uncertainties, it could happen many years later, or could happen as early as 2023.

Not all sources will have associated transient electromagnetic counterparts: we may fail to identify the counterparts to some BNS mergers, and counterparts are not expected for BBH mergers. For cases where a unique counterpart cannot be identified, it is possible to carry out a measurement of the Hubble constant using the statistical approach.

To do this, the redshifts of all potential host galaxies within the gravitational-wave three-dimensional localization region are incorporated, yielding an $H_0$ measurement that is inferior to what can be calculated using a counterpart, but is still informative once many detections are combined. This means that, in the absence of a counterpart, only those gravitational-wave events with small enough localization volumes yield informative $H_0$ measurements. If the localization volume is too large, it contains a large number of potential host galaxies, which will largely wash out the contribution from the correct host galaxy. Additionally, it may be difficult in practice to construct a complete galaxy catalogue over a large volume with precise galaxy redshifts. We find that for BBHs without counterparts, combining the $H_0$ measurement from events that are localized to within 10,000 Mpc$^3$ (approximately 40% of events) yields identical constraints to the combined measurement using the full sample—events localized to greater than 10,000 Mpc$^3$ do not contribute to the measurement. For this reason, we use only the sources localized to within 10,000 Mpc$^3$ for the no-counterpart projections in Fig. 2. We note that for all of the no-counterpart curves in Fig. 2, we start with a flat $H_0$ prior in the range 50–100 km s$^{-1}$ Mpc$^{-1}$.

Because BBH systems tend to have much larger localization volumes than BNS systems (as they are more massive and found at greater distances), the statistical $H_0$ measurement for BBHs converges very slowly, even though they are detected at higher rates. We consider both 'light' (components of mass $10M_\odot$, where $M_\odot$ is the solar mass; denoted as '10M$_\odot$–10M$_\odot$') and 'heavy' (components of mass 30M$_\odot$; denoted as 30M$_\odot$–30M$_\odot$) BBHs, assuming merger rates of $8^{+90}_{-70}$ Gpc$^{-3}$ yr$^{-1}$ for the 10M$_\odot$–10M$_\odot$ BBHs and 11$^{+11}_{-9}$ Gpc$^{-3}$ yr$^{-1}$ for the 30M$_\odot$–30M$_\odot$ BBHs (see Methods). Only about 3% of the bright BBHs are localized to within 10,000 Mpc$^3$, which means that we expect to detect only 16$^{+25}_{-16}$ well localized BBHs by 2026. This leads to an approximately 10% $H_0$ measurement with BBHs by 2026. We note that the constraints from statistical BBH standard sirens improve if the BBH rates are on the high end, as well as if the BBH mass function favours low masses.

For the projections in Fig. 2, we assumed that galaxies are distributed uniformly in a comoving volume and that complete catalogues are available. If we incorporate the clustering of galaxies due to large-scale structure, the convergence rate in the statistical case improves by a factor of about 2.5 (see Methods). Incorporating this large-scale structure effect, we find that we will still need to detect more than about 50 BNSs without a counterpart to reach a 6% $H_0$ measurement, compared to only ten BNSs or fewer with a counterpart. Meanwhile, accounting for galaxy catalogue incompleteness provides an additional source of uncertainty (see equation (5) in Methods), which can cancel out some of the improvement due to large-scale structure. For example, for a galaxy catalogue completeness of 50%, the $H_0$ measurement would be degraded by about a factor of two. Therefore, incorporating the effects of large-scale structure and catalogue incompleteness, we expect that in practice the $H_0$ constraints in the statistical case will be slightly better than our prediction in Fig. 2, where the precise factor depends on properties of the relevant host galaxies and completeness of the catalogue.

Besides the with-counterpart and without-counterpart cases, we can also anticipate a situation in which we have a counterpart detection but no unambiguous host association. For example, an optical counterpart could be relatively isolated on the sky without a clearly identified host galaxy, or may have multiple possible host galaxies. In this case we can pursue a pencil-beam strategy, for example, focusing on the volume within 100 kpc of the counterpart (see Methods). For BNSs this will reduce the relevant volume to about 10 Mpc$^3$, for which we expect to have only about one potential host galaxy or galaxy group, which thereby reduces to the with-counterpart case.

In addition to the BNSs and BBHs discussed here, we can expect mergers of a neutron star and a black hole$^{19–21}$, and these may have detectable
electromagnetic counterparts. Although the rates for these systems are uncertain and expected to be low, they will also be seen to greater distances than BNS systems, which may render them useful as standard sirens.22

We note that our measurements of distance do not use any astrophysical modelling. However, associated electromagnetic observations (for example, from short γ-ray burst afterglows or jet breaks) can provide additional constraints on the inclination, and thereby improve the individual measurements22,23 of distance, leading to a tighter measurement of $H_0$. In this sense our counterpart results can be considered a conservative estimate. However, one of the advantages of standard sirens is that they are 'pure' measurements of luminosity distance, avoiding complicated astrophysical distance ladders or poorly understood calibration processes, and instead are calibrated directly by the theory of general relativity to cosmological distances. By introducing additional constraints based on astronomical observations (for example, independent beaming measurements or estimates of the mass distribution or equation of state of neutron stars), there is the potential to introduce systematic biases that could fundamentally contaminate the standard siren measurements. In the present analysis we do not consider these additional constraints, although they may indeed have an important part to play in future standard siren science.

Eventually systematic errors in the amplitude calibration of the detectors may become a source of concern, because the luminosity distance is encoded in the amplitude of the gravitational-wave signal. However, the calibration uncertainty is currently limited by the photon calorimeter to around 1%, and this is likely to improve; we look forward to an era where sub-1% calibration becomes a necessity, but this is a number of years away. Another possible source of distance uncertainty is gravitational lensing. However, at the typical redshift of BNS and BBH systems ($z < 0.5$ at design sensitivity) the effect will be minor relative to the uncertainty from the distance–inclination degeneracy.25 In addition, for sufficient numbers of sources the effects of lensing will average away.26 Of course, gravitational-wave cosmology is a new field, and unforeseen systematics could certainly arise as we push our measurements to the 1% level and beyond.

We stress that our projected $H_0$ constraints are subject to several important uncertainties, the largest one of which is the merger rate of BNS and BBM systems. The detection rate for BBHs depends sensitively on the mass distribution, which is not currently well constrained.24–27 Future detections will bring a better understanding of the merger rates and mass distributions of compact objects, allowing for improved predictions. Nevertheless, it is clear that gravitational-wave standard sirens will provide precision constraints on cosmology in the upcoming advanced-detector era of gravitational-wave astronomy.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-018-0606-0.

Received: 9 January; Accepted: 16 August 2018; Published online 17 October 2018.

1. Schutz, B. F. Determining the Hubble constant from gravitational-wave observations. *Nature* **323**, 310–311 (1986).
2. Holz, D. E. & Hughes, S. A. Using gravitational-wave standard sirens. *Astrophys. J.* **629**, 15 (2005).
3. Dalal, N., Holz, D. E., Hughes, S. A. & Jain, B. Short GRB and binary black hole standard sirens as a probe of dark energy. *Phys. Rev. D* **74**, 063006 (2006).
4. Abbott, B. P. et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature* **551**, 85–88 (2017).
5. Coulter, D. A. et al. Swope Supernova Survey 2017a (SSS17a), the optical counterpart to a gravitational wave source. *Science* **358**, 1556–1558 (2017).
6. Soares-Santos, M. et al. The electromagnetic counterpart of the binary neutron star merger LIGO/Virgo GW170817. I. Discovery of the optical counterpart using the dark energy camera. *Astrophys. J.* **848**, L16 (2017).
7. Riess, A. G. et al. A 2.4% determination of the local value of the Hubble constant. *Astrophys. J.* **826**, 56 (2016).
8. Aghanim, N. et al. Planck 2018 results. VI. Cosmological parameters. Preprint at https://arxiv.org/abs/1807.06209 (2018).
9. Riess, A. G. et al. New paradigms of galactic Cepheids from spatially scanning the Hubble Space Telescope: implications for the Hubble constant. *Astrophys. J.* **855**, 136 (2018).
10. Hu, W. Dark energy probes in light of the CMB. *ASP Conf. Series* **339**, 215 (2005); preprint at https://arxiv.org/abs/astro-ph/0407158.
11. Di Valentino, E., Holz, D. E., Melchiorri, A. & Renzi, F. The cosmological impact of future constraints on $H_0$ from gravitational-wave standard sirens. Preprint at https://arxiv.org/abs/1806.07463 (2018).
12. Li, L.-X. & Paczynski, B. Transient events from neutron star mergers. *Astrophys. J.* **507**, L59–L62 (1998).
13. Metzger, B. D. et al. Electromagnetic counterparts of compact object mergers powered by the radioactive decay of $r$-process nuclei. *Mon. Not. R. Astron. Soc.* **408**, 2650–2662 (2010).
14. Feng, W. & Berger, E. The locations of short gamma-ray bursts as evidence for compact object binary progenitors. *Astrophys. J.* **776**, 18 (2013).
15. Hirata, C. M., Holz, D. E. & Cutler, C. Reducing the weak lensing noise for the gravitational wave Hubble diagram using the non-Gaussianity of the magnification distribution. *Phys. Rev. D* **81**, 124046 (2010).
16. Nissanka, S. et al. Determining the Hubble constant from gravitational wave observations of merging compact binaries. Preprint at https://arxiv.org/abs/1307.2638 (2013).
17. Abbott, B. P. et al. GW170817: observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.* **119**, 161101 (2017).
18. Abbott, B. P. et al. GW170104: observation of a 50-solar-mass binary black hole coalescence at redshift 0.2. *Phys. Rev. Lett.* **118**, 221101 (2017).
19. Dominik, M. et al. Double compact objects III: gravitational-wave detection rates. *Astrophys. J.* **806**, 263 (2015).
20. Belczynski, K. et al. Compact binary merger rates: comparison with LIGO/Virgo upper limits. *Astrophys. J.* **819**, 108 (2016).
21. Belczynski, K., Holz, D. E., Bulik, T. & O’Shaughnessy, R. The first gravitational-wave source from the isolated evolution of two stars in the 40–100 solar mass range. *Nature* **534**, 512–515 (2016).
22. Vitale, S. & Chen, H.-Y. Measuring the Hubble constant with neutron star black hole mergers, *Phys. Rev. Lett.* **121**, 021303 (2018).
23. Guidorzi, C. et al. Improved constraints on $H_0$ from a combined analysis of gravitational-wave and electromagnetic emission from GW170817. *Astrophys. J.* **851**, L26 (2017).
24. Karki, S. et al. The Advanced LIGO photon calibrators. *Rev. Sci. Instrum.* **87**, 114503 (2016).
25. Holz, D. E. & Linder, E. V. Safety in numbers: gravitational lensing degradation of the luminosity distance-redshift relation. *Astrophys. J.* **631**, 678–686 (2005).
26. Abbott, B. P. et al. The rate of binary black hole mergers inferred from advanced LIGO observations surrounding GW150914. *Astrophys. J.* **833**, 11 (2016).
27. Fishbach, M. & Holz, D. E. Where are LIGO’s big black holes? *Astrophys. J.* **851**, L29 (2017).

Acknowledgements We acknowledge discussions with L. Blackburn, R. Essick, W. Farr and J. Gair. We were supported in part by NSF CAREER grant PHY-1151836 and NSF grant PHY-1708081. We were also supported by the Kavli Institute for Cosmological Physics at the University of Chicago through NSF grant PHY-1125897 and an endowment from the Kavli Foundation. We acknowledge the University of Chicago Research Computing Center for support of this work. H.-Y.C. was supported in part by the Black Hole Initiative at Harvard University, through a grant from the John Templeton Foundation. M.F. was supported by the NSF Graduate Research Fellowship Program under grant DGE-1746045.

Author contributions H.-Y.C. led the project, conducted the simulations and led the analysis. M.F. provided the mathematical derivations and contributed to the analysis and results. D.E.H. conceived the project, supervised the research, and contributed to the analysis and results. All authors contributed to the draft preparation.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-018-0606-0.

Reprints and permissions information is available at http://www.nature.com/reprints.

Correspondence and requests for materials should be addressed to H.-Y.C.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
METHODS

We present our method for inferring cosmological parameters from gravitational-wave (GW) and electromagnetic (EM) measurements. We first simulate a representative sample of GW detections for a range of detector configurations. We then simulate the analysis of these data sets, and explore the resulting standard sirens constraints. In what follows we highlight important aspects of our calculation, such as the role of peculiar velocities and selection effects.

Syntetically generated host galaxies. Measuring $H_0$ with standard sirens relies on our ability to extract the luminosity distance and sky position of GW sources. We follow the procedure in ref. 28 to localize synthetic BNS merger and BBH merger detections. The population of binaries are distributed uniformly in comoving volume in a Planck (2015) cosmology ($\Omega_m = 0.308$, $\Omega_{\Lambda} = 0.692$, $H_0 = 67.8$). We assume that the BNS merger rate follows the rate measured in ref. 17. To estimate the merger rate of $10^{-5}$–$10^{-3}$ and $30 M_\odot$–$300 M_\odot$ BBHs from the rate measured in ref. 18, we assume that the BBH mass function follows a Salpeter power law and use $10 M_\odot$–$100 M_\odot$ BBHs to characterize all BBHs with primary component masses between $5 M_\odot$ and $15 M_\odot$ and $30 M_\odot$–$300 M_\odot$ BBHs to characterize all BBHs with primary component masses between $20 M_\odot$ and $50 M_\odot$. We do not place additional cuts on the secondary masses, which are distributed uniformly between $5 M_\odot$ and the primary mass.

The detection rate of sources depends on the sensitivity, observing time and duty cycle of the GW detector network. We assume that the LIGO–Virgo network operates for one year at projected O3 sensitivity, followed by two one-year-long observing runs of LIGO-Hanford + LIGO-Livingston + Virgo (HLV) at design sensitivity and two one-year-long runs of the five-detector network, LIGO-Hanford + LIGO-Livingston + Virgo + LIGO-India (HLVII), at design sensitivity. We take the combined duty cycle to be 0.8 for the HLV detector configuration and 0.3 for HLVII. The number of detections is subject to Poisson statistics, and we simulate detections according to the merger rate, network sensitivity, observing time and duty cycle.

To determine whether a binary merger is detected, we calculate the matched-filter signal-to-noise ratio (SNR) for each simulated binary. We draw the ‘measured’ SNR from a Gaussian distribution centred at the matched-filter value with a standard deviation of $\sigma = 1$. Binary mergers are detected only if their measured network SNR is greater than 12. For each detected merger, we calculate its three-dimensional localization according to the methods in ref. 28. (We have verified that this procedure yields results which are consistent with the full parameter estimation pipeline, LALInference.) The three-dimensional localization takes the form of a posterior probability distribution function, $p(\alpha, \delta, D_{\text{gw}} | d_{\text{gw}})$, of the sky position $(\alpha, \delta)$ and luminosity distance, $D_L$, given the GW data, $d_{\text{gw}}$. The GW signal from each detected binary merger provides a measurement of $D_L$. To calculate $H_0$, we must also measure a redshift for each binary merger. Throughout, we take the redshift, $z$, to be the peculiar-velocity-corrected redshift: that is, the redshift one would have if it were measured in the Hubble flow. We consider two cases: the redshift information either comes from a direct EM counterpart, such as a short γ-ray burst/afterglow and/or a kilonova (‘with-counterpart’), or a statistical analysis over a catalogue of potential host galaxies (‘statistical’).

In the with-counterpart case, we assume that the EM counterpart is close enough to its host galaxy that the host can be unambiguously identified, and we can measure its sky position and redshift. This is a reasonable assumption based on the distribution of offsets between short γ-ray bursts and their host galaxies, assuming that short γ-ray bursts trace a population similar to that of BNS mergers, and taking into account that detected BNS mergers will be at much lower redshifts than the short γ-ray burst population. We assume that the sky position of each host galaxy is perfectly measured (that is, with negligible measurement error), meaning we can fix the source sky position to the location of the counterpart in the GW parameter estimation (rather than marginalizing over all sky positions). The GW horizon distance changes slowly over the sky and therefore is not sensitive to the precise location of the counterpart. However, since the GW sky localization areas can be very large, fixing the source position can lead to important improvements in the distance, and hence $H_0$, measurements. We also assume that the peculiar-velocity-corrected redshift, $z$, is measured with a 1σ error of 200 km s$^{-1}$, where $c$ is the speed of light in vacuum, which is a typical uncertainty for the peculiar velocity correction.

In the absence of an EM counterpart we cannot identify a single host galaxy, and must use a catalogue of all potential host galaxies. To simulate the galaxy catalogues we consider two cases: a uniform-in-comoving-volume distribution of galaxies, and a distribution that follows the large-scale structure as simulated by the MICE galaxy catalogue. In the uniform distribution case, we construct a mock catalogue by distributing galaxies uniformly in comoving volume with a number density of $1.6 \times 10^{-5}$ Mpc$^{-3}$. This corresponds to the number density of galaxies that are $25\%$ as bright as the Milky Way, assuming the galaxy luminosity function is described by the Schechter function with $B$-band parameters $\alpha_{\text{S}} = 1.6 \times 10^{-5}$ Mpc$^{-3}$, $\alpha_{11} = -1.07$, $L_{\text{B}}^* = 1.2 \times 10^{10} L_\odot$, and $h = 0.7$ (where $L_{\text{B}}^*$ is the solar luminosity in the B band), and integrating down to $0.16 L_{\text{B}}^*$ to find the luminosity density. (This corresponds to $83\%$ of the total luminosity.) The lower luminosity limit of the MICE catalogue is similar. Thus, we assume that only galaxies brighter than $0.16 L_{\text{B}}^*$ can host binary mergers, although we note that the population of host galaxies is currently uncertain, and we can modify the assumed luminosity limit by including the effects of catalogue incompleteness. A lower luminosity limit would increase the galaxy density and weaken the $H_0$ constraints in the $H_0$ uncertainty. Not all GW events contribute equally to the $H_0$ measurements. In the counterpart case, the fractional error on the $H_0$ measurement from a single source depends on the fractional distance uncertainty of the GW source and the fractional redshift uncertainty of its host galaxy. To first order, this is:

$$\frac{\sigma_{\text{GW}}}{H_0} \approx \frac{\sigma_{\text{ref}}}{H_0} \left(1 + \frac{\sigma_{\text{ref}}}{H_0} \right)^2$$

where ‘1gal’ denotes the case of a uniquely identified host galaxy and $\nu_{\text{ref}}$ is the peculiar-velocity-corrected ‘Hubble velocity’. Because the recessional velocity uncertainty, $\sigma_{\text{ref}}$, is typically around $150$–$250$ km s$^{-1}$, the fractional recessional velocity decreases with distance. Meanwhile, the fractional distance uncertainty scales roughly inversely with SNR, and therefore tends to increase with distance. There is thus a sweet spot, at which the peculiar velocities and the distance uncertainties are comparable; for LIGO–Virgo’s second observing run, this was about 30 Mpc, near the distance of GW170817. The distance of the sweet spot will increase as the networks become more sensitive; for detectors at design sensitivity the ideal BNS standard siren distance will be about 50 Mpc. At distances beyond this, the distance uncertainty will tend to dominate the peculiar velocity uncertainty; in this regime, the nearest (highest SNR) events tend to provide the tightest $H_0$ constraints. This can be seen in Extended Data Fig. 1, which shows the fractional $H_0$ uncertainty for individual events, plotted against the median posterior distance and $90\%$ posterior localization volumes. However, we note that the relationship between median distance, localization volumes, and fractional $H_0$ uncertainty is not very tight. Prior to identifying the counterpart for a particular event, we can estimate the accuracy of the $H_0$ measurement from the width and central value (for example, median) of the GW distance posterior according to equation (1), using an estimated $\nu_{\text{ref}} \approx 70(42)\text{ km s}^{-1}$ Mpc$^{-1}$, where $(D_L)$ is the median GW distance. (Here we must use the GW posterior marginalized over the sky position, as we do not yet know the sky position of the counterpart.) We verify that this estimate of the combined distance and redshift uncertainty is a reasonable proxy for the resulting $H_0$ uncertainty, assuming that an EM counterpart is found and provides an independent measurement of redshift.

In the absence of a counterpart, we cannot assign a unique host, and so the $H_0$ error increases with the number of potential host galaxies in the localization volume. Galaxy clustering can mitigate this, as we discuss in the main text. For example, in the case of GW170817, the optical counterpart was found in NGC 4993, which is a member of a group of about 20 galaxies, all of which have an equivalent Hubble recessional velocity$^{36}$. On the other hand, catalogue incompleteness degrades the $H_0$ measurement, as we have to consider an additional background of uniformly distributed galaxies (see equation (5)).

Bayesian model. For a single event with GW and EM data, $d_{\text{gw}}$ and $d_{\text{EM}}$, we can write the likelihood as:

$$p(d_{\text{gw}}, d_{\text{EM}} | H_0) = \int p(d_{\text{gw}} | d_{\text{EM}}, D_L, \alpha, \delta, z | H_0) p(D_L) dD_L d\alpha d\delta dz \approx p(d_{\text{gw}} | d_{\text{EM}}, H_0)$$

where we have included a normalization term in the denominator, $p(H_0)$, to account for selection effects and ensure that the likelihood integrates to unity. We can factor the numerator in equation (2) as:

$$\int p(d_{\text{gw}} | d_{\text{EM}}, D_L, \alpha, \delta, z | H_0) dD_L d\alpha d\delta dz$$

$$= \int p(d_{\text{gw}} | D_L, \alpha, \delta) p(D_L | z, H_0) p(z, \alpha, \delta | H_0) dD_L d\alpha d\delta dz$$

$$= \int p(d_{\text{gw}} | D_L, \alpha, \delta) p(D_L | z, H_0) p(z, \alpha, \delta | H_0) dD_L d\alpha d\delta dz$$

$$= \int p(d_{\text{gw}} | D_L, \alpha, \delta) p(D_L | z, H_0) p(z, \alpha, \delta | H_0) dD_L d\alpha d\delta dz$$

where $D_L(z, H_0)$ denotes the luminosity distance of a source at redshift $z$, given a Hubble constant of $H_0$ and leaving all other cosmological parameters fixed to the Planck values$^{36}$ ($\Omega_m = 0.308$, $\Omega_{\Lambda} = 0.692$). We can, alternatively, marginalize...
over these other cosmological parameters, but since most detected binaries will be at low redshifts, the effects of other cosmological parameters on the $z - D_L$ relation are small. The term $\rho(d_{GW}|D_L, \alpha, \delta)$ is the marginalized likelihood of the GW data given a compact binary source at distance $D_L$ and sky position $(\alpha, \delta)$, marginalized over all other parameters. Throughout, we assume that we can construct a catalogue of the potential host galaxies for each event, and take the prior $p(z, \sigma, D_L|H_0)$ to be a sum of Gaussian distributions centred at the measured redshifts and sky positions of the galaxies:

$$p(z, \sigma, D_L|H_0) = p_{\text{catalogue}}(z, \sigma, D_L) = \frac{1}{N_{\text{gal}}} \sum_{i=1}^{N_{\text{gal}}} N(\sigma_i, \sigma_0|z_i, \sigma_i|D_L_i)$$

where $p_{\text{catalogue}}$ is given by equation (4), and:

$$p_{\text{max}}(z, \sigma, D_L|H_0) \propto \left\{1 - P_{\text{complex}}(z, \sigma, D_L)\right\} \frac{dV}{dz d\delta}$$

where $P_{\text{complex}}$ is the probability of a galaxy at $(z, \sigma, D_L)$ being in the catalogue. Meanwhile the completeness fraction is given by:

$$f = \frac{1}{V(z_{\text{max}})} \int_0^{z_{\text{max}}} \int_0^{D_L(z)} P_{\text{complex}}(z, \sigma, D_L) \frac{dV}{dz d\delta}$$

where $z_{\text{max}}$ is the maximum galaxy redshift considered in the analysis of an individual event, and $V(z_{\text{max}})$ is the total comoving volume enclosed within $z_{\text{max}}$.

In the case where we have an EM counterpart, the likelihood $p(d_{EM}|z, \sigma, D_L)$ picks out one of the galaxies in the catalogue, so that the sum in the prior reduces to a single term corresponding to the EM-identified host galaxy. In the case where there is no EM counterpart, the EM data are uninformative, and we set the likelihood $p(d_{EM}|z, \sigma, D_L) \propto constant$. In the case where we have an EM counterpart but cannot pick out a unique host galaxy, one could consider a ‘pencil beam’ containing all the potential host galaxies within around 100 kpc in projected distance on the sky. We assume the sky position of the counterpart is perfectly measured to be $(\sigma, \delta)$, and take the term $p(d_{EM}|z, \sigma, D_L)$ to be a top hat that picks out all of the galaxies within some angular radius $r$ (corresponding to around 100 kpc in projected distance) of the counterpart’s sky position. Thus, the numerator of equation (3) reduces to:

$$\int_{\{(\sigma, \delta)|D_L(z) < D_L\}} p(d_{GW}|D_L(z), \sigma, \delta) d\sigma d\delta d\delta$$

and we sum over all galaxies within about 100 kpc in projected distance, but no longer weight them by the likelihood of the GW source at the corresponding sky position. Alternatively, we can incorporate assumptions about the kick distribution in the form of $p(d_{GW}|z, \sigma, D_L)$ and place more weight at galaxies close to $(\sigma, \delta)$ rather than assuming a simple top hat. Although for simplicity we do not apply the pencil-beam approach in this Letter, it can be thought of as a natural interpolation between the counterpart and statistical cases.

To calculate the normalization term, $\beta(H_0)$, in the denominator of equation (2), we must account for selection effects in our measurement process. In general, the GW and EM data are both subject to selection effects in that we only detect GW and EM sources that are above some threshold, $d_{GW}^{\text{thr}}$ and $d_{EM}^{\text{thr}}$, respectively. Accounting for these detection thresholds, the denominator of equation (2) is:

$$\beta(H_0) = \frac{1}{N_{\text{gal}}} \sum_{i=1}^{N_{\text{gal}}} \int_{d_{GW}^{\text{thr}}}^{d_{GW}^{\text{thr}}} \int_{D_L^{\text{thr}}}^{\infty} \int_{d_{EM}^{\text{thr}}}^{d_{EM}^{\text{thr}}} p(d_{GW}|D_L, \sigma, \delta, z) d\sigma d\delta d\sigma d\delta d\delta d\delta d\sigma d\delta d\delta$$

We define:

$$P_{\text{det}}^{GW}(D_L, \sigma, \delta, z) \equiv \int_{d_{GW}^{\text{thr}}}^{d_{GW}^{\text{thr}}} \int_{D_L^{\text{thr}}}^{\infty} \int_{d_{EM}^{\text{thr}}}^{d_{EM}^{\text{thr}}} p(d_{GW}|D_L, \sigma, \delta, z) d\sigma d\delta d\sigma d\delta d\delta d\delta d\sigma d\delta d\delta$$

and similarly:

$$P_{\text{det}}^{EM}(z, \sigma, \delta) \equiv \int_{d_{EM}^{\text{thr}}}^{d_{EM}^{\text{thr}}} \int \int p(d_{EM}|z, \sigma, \delta) d\sigma d\delta d\sigma d\delta d\delta d\delta d\sigma d\delta d\delta$$

With these definitions, equation (9) becomes:

$$\beta(H_0) = \int \int P_{\text{det}}^{GW}(D_L(z, \sigma, \delta), \sigma, \delta) \rho(z) d\sigma d\delta d\sigma d\delta d\sigma d\delta d\sigma d\delta d\sigma$$

Note that we have applied the same chain-rule factorization to the inner four integrals as in equation (3). It is clear that the normalization factor $\beta(H_0)$ depends on $H_0$, so it is crucial to include it in the likelihood. For the EM selection effects, we assume that we can detect all EM counterparts and host galaxies up to some maximum true redshift, $z_{\text{max}}$. This is an over-simplification of the true EM selection effects, but is a reasonable assumption for the real-time galaxy catalogues that will be constructed during the EM follow-up to GW events. For example, at Advanced LIGO design sensitivity, 90% of 30 Mpc–300 Mpc BBH detections will be within 5 Gpc (and BNS detections will be within 0.3 Gpc) (see http://gwc.rcc.uchicago.edu/). Furthermore, only the BBHs with the smallest localization volumes contribute to the $H_0$ constraints, and these will typically be within 400 Mpc. A galaxy with the same absolute magnitude as the Milky Way would have an apparent magnitude of $-23$ at typical 30 Mpc–300 Mpc BBH distances, or $-17.5$ for well localized BBHs, and $<17$ at typical BNS distances. Meanwhile, we expect kilonova counterparts to BNS mergers to have magnitudes of $\leq 21.7$ on the first night, even at the farthest distances detectable by the HLMT network at design sensitivity (assuming the optical counterpart to GW170817 is typical (14)), this is well within the magnitude limits of upcoming survey telescopes (for example, the Large Synoptic Survey Telescope), as well as within reach of current instruments (such as the Dark Energy Camera, the Subaru Hyper Suprime-Cam, the Zwicky Transient Facility and so on).

We assume that EM counterparts are detectable for binaries regardless of the binary inclination. Although the short gravitational burst duration is expected to be beamed, the associated kilonovae are expected to emit isotropically. Furthermore, as GW170817 demonstrated, it is possible to identify a kilonova counterpart independently of the gravitational burst (see the discovery of the optical counterpart to GW170817) (14). This is well within the magnitude limits of upcoming survey telescopes (for example, the Large Synoptic Survey Telescope), as well as within reach of current instruments (such as the Dark Energy Camera, the Subaru Hyper Suprime-Cam, the Zwicky Transient Facility and so on).

We assume that EM counterparts are detectable for binaries regardless of the binary inclination. Although the short gravitational burst duration is expected to be beamed, the associated kilonovae are expected to emit isotropically. Furthermore, as GW170817 demonstrated, it is possible to identify a kilonova counterpart independently of the gravitational burst (see the discovery of the optical counterpart to GW170817) (14). This is well within the magnitude limits of upcoming survey telescopes (for example, the Large Synoptic Survey Telescope), as well as within reach of current instruments (such as the Dark Energy Camera, the Subaru Hyper Suprime-Cam, the Zwicky Transient Facility and so on).

We assume that EM counterparts are detectable for binaries regardless of the binary inclination. Although the short gravitational burst duration is expected to be beamed, the associated kilonovae are expected to emit isotropically. Furthermore, as GW170817 demonstrated, it is possible to identify a kilonova counterpart independently of the gravitational burst (see the discovery of the optical counterpart to GW170817) (14). This is well within the magnitude limits of upcoming survey telescopes (for example, the Large Synoptic Survey Telescope), as well as within reach of current instruments (such as the Dark Energy Camera, the Subaru Hyper Suprime-Cam, the Zwicky Transient Facility and so on).
We note that for GW sources in the local Universe ($D_L < 50$ Mpc):

$$D_L(z, H_0) \approx c z / H_0$$

(16)

If we assume that the distribution of galaxies is uniform in comoving volume, then in the local universe, we can approximate:

$$p_z(z) \propto z^2$$

(17)

With these approximations, assuming that EM selection effects are negligible ($z_{\text{max}} \to \infty$), $\beta(H_0)$ is independent of the masses of the source, which determine the distance to which it can be detected. In fact, under these assumptions, $\beta(H_0)$ simplifies to:

$$\beta(H_0) \propto H_0^3$$

(18)

However, in general, we must account for cosmological deviations from equations (16) and (17), so we calculate $\beta(H_0)$ according to equation (15) throughout our analysis. We note that $\beta(H_0)$ is still only weakly dependent on the GW horizon and therefore on the (unknown) underlying mass distribution of GW sources. Nevertheless, the statistical framework described here can accommodate more complicated models of the GW source distribution and its effects on the detection probability (equation (14)).

**Data availability**

Source Data for Figs. 1, 2 and Extended Data Fig. 1 are provided with the online version of the paper. Other data that support the findings of this study are available from the corresponding author upon reasonable request.

28. Chen, H.-Y. & Holz, D. E. Finding the one: identifying the host galaxies of gravitational-wave sources. Preprint at https://arxiv.org/abs/1612.01471 (2016).

29. Abbott, B. P. et al. Prospects for observing and localizing gravitational-wave transients with advanced LIGO and advanced Virgo. *Liv. Rev. Rel.* **19**, 1 (2016).

30. Veitch, J. et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D* **91**, 042003 (2015).

31. Carrick, J., Turnbull, S. J., Lavaux, G. & Hudson, M. J. Cosmological parameters from the comparison of peculiar velocities with predictions from the 2M++ density field. *Mon. Not. R. Astron. Soc.* **450**, 317–332 (2015).

32. Scolnic, D. M. et al. The complete light-curve sample of spectroscopically confirmed type Ia supernovae from Pan-STARRS1 and cosmological constraints from the combined Pantheon sample. *Astrophys. J.* **859**, 101 (2018).

33. Del Pozzo, W. Inference of cosmological parameters from gravitational waves: applications to second generation interferometers. *Phys. Rev. D* **86**, 043011 (2012).

34. Fosalba, P., Croce, M., Gaztañaga, E. & Castander, F. J. The MICE grand challenge lightcone simulation—I. Dark matter clustering. *Mon. Not. R. Astron. Soc.* **448**, 2987–3000 (2015).

35. Croce, M., Castander, F. J., Gaztañaga, E., Fosalba, P. & Carretero, J. The MICE Grand Challenge lightcone simulation—II. Halo and galaxy catalogues. *Mon. Not. R. Astron. Soc.* **453**, 1513–1530 (2015).

36. Fosalba, P., Gaztañaga, E., Castander, F. J. & Croce, M. The MICE Grand Challenge light-cone simulation—III. Galaxy lensing mocks from all-sky lensing maps. *Mon. Not. R. Astron. Soc.* **447**, 1319–1332 (2015).

37. Carretero, J. et al. CoSMoHub and SciPIC: massive cosmological data analysis, distribution and generation using a Big Data platform. In *Proc. 2017 European Physical Society Conference on High Energy Physics* 488 (EPS-HEP, 2017).

38. Schechter, P. An analytic expression for the luminosity function for galaxies. *Astrophys. J.* **203**, 297–306 (1976).

39. Norberg, P. et al. The 2dF Galaxy Redshift Survey: the $b$-band galaxy luminosity function and survey selection function. *Mon. Not. R. Astron. Soc.* **336**, 907–931 (2002).

40. Liske, J., Lemon, D. J., Driver, S. P., Cross, N. J. G. & Couch, W. J. The Millennium Galaxy Catalogue: 16 $\leq B_{\text{MGC}} < 24$ galaxy counts and the calibration of the local galaxy luminosity function. *Mon. Not. R. Astron. Soc.* **344**, 307–324 (2003).

41. González, R. E., Lares, M., Lambas, D. G. & Valotto, C. The faint-end of the galaxy luminosity function in groups. *Astron. Astrophys.* **445**, 51–58 (2006).

42. Kourkchi, E. & Tully, R. B. Galaxy groups within 3500 km s$^{-1}$. *Astrophys. J.* **843**, 16 (2017).

43. Chen, H.-Y. et al. Distance measures in gravitational-wave astrophysics and cosmology. Preprint at https://arxiv.org/abs/1709.08079 (2017).

44. Abbott, B. P. et al. Multi-messenger observations of a binary neutron star merger. *Astrophys. J.* **848**, 12 (2017).

45. Schutz, B. F. Networks of gravitational wave detectors and three figures of merit. *Class. Quant. Gravity* **28**, 125023 (2011).

46. Ade, P. A. R. et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.* **594**, A13 (2016).
Extended Data Fig. 1 | $H_0$ uncertainty for BNS systems with identified counterparts and redshifts. Each point is the $H_0$ uncertainty $\sigma_{H_0}$ from a simulated detection with the Advanced HLV network operating at design sensitivity, as a function of the 90% localization volume. The colours correspond to the median of the GW distance measurement. The lower limit to the precision of individual measurements of about $3 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is due to the ‘sweet spot’ between peculiar velocities and distance uncertainties, as discussed in the text. We find that, in general, closer events have smaller localization volumes and lead to better constraints on $H_0$, although the closest events yield slightly worse constraints because of their larger fractional peculiar velocity uncertainties.